

Gencore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using fframe_plus.p2n model

Run on: February 26, 2003, 19:53:56 ; Search time 3193 Seconds

(without alignments)
5897.123 Million cell updates/sec

Title: US-09-677-653a-50
Perfect score: 3374
Sequence: 1 MDDAGVASRPNHRRCTRN.....GRIARVRRARRRRARRAN 647

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=fringe+.p2n.model -DEV=xlh
-Q=/cgn2.1/USPTO.spool/US09677653/runat_20022003_164315_1251/app_query.fasta.1.839
-DB=Gencore -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blousum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09677653.ecgn.1.1_2034.etrnat_20022003_164315_1251 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Gencore: *
1: gb_da: *
2: gb_htg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_dat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_da: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *

29: em_vl: *
30: em_htg_hum: *
31: em_htg_inv: *
32: em_htg_other: *
33: em_htg_mus: *
34: em_htg_pln: *
35: em_htg_rod: *
36: em_htg_mam: *
37: em_htg_vrt: *
38: em_sy: *
39: em_htgo_hum: *
40: em_htgo_mus: *
41: em_htgo_other: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3374	100.0	2478	6 AR125096	AR125096 Sequence
2	3374	100.0	2478	6 AR125097	AR125097 Sequence
3	3374	100.0	2478	6 AR159968	AR159968 Sequence
4	3374	100.0	2478	6 AX028950	AX028950 Sequence
5	3374	100.0	2478	6 AX028971	AX028971 Sequence
6	3374	100.0	2478	14 HVS0098	L37299 Helicoverpa
7	3361	99.6	2479	6 AR125098	AR125098 Sequence
8	2207	65.4	2448	14 AF062037	S43937 capsid prot
9	436	12.9	2482	14 AF062037	AF062037 Thosaa as
10	407.5	12.1	5638	14 AF461742	AF461742 Euproster
11	386.5	11.5	6625	14 AF102884	AF102884 Nudaurell
12	294	8.7	6534	6 AR159969	AR159969 Sequence
13	294	8.7	6536	6 AX028953	AX028953 Sequence
14	294	8.7	6536	6 AX028972	AX028972 Sequence
15	155.5	4.6	11659	1 AE005131	AE005131 Halobacte
16	153	4.5	151826	2 AC125376	AC125376 Mus muscu
17	148.5	4.4	110000	2 LMFICHR31_05	Continuation (6 of
18	141.5	4.2	2663	14 EBYBLLE1	X99106 Epstein-Bar
19	140.5	4.2	8516	1 AE004995	AE004995 Pseudomon
20	140.5	4.2	11242	1 AE004995	AE004995 Halobacte
21	140.5	4.2	12041	1 AE0010954	AE0010954 Methanosa
22	140	4.1	161563	8 AP003076	AP003076 Oryza sat
23	138.5	4.1	139517	9 AC105446	AC105446 Homo sapi
24	138	4.1	4974	3 S83394	S83394 Insulin-lik
25	137	4.1	11482	1 AE001876	AE001876 Deinococc
26	136.5	4.0	27887	1 SCE50	AL163672 Streptococ
27	136	4.0	37620	1 SC1A6	AL023496 Streptomy
28	136	4.0	310050	1 RME603642	AL603642 Rhizobium
29	135	4.0	1891	1 SBSTRBF	X78972 S. bluenis
30	134.5	4.0	2535	9 AF152475	AF152475 Homo sapi
31	134.5	4.0	5254	9 AF152306	AF152306 Homo sapi
32	134.5	4.0	15742	1 AF013216	AF013216 Myxococcu
33	134.5	4.0	157970	9 AC005609	AC005609 Homo sapi
34	134	4.0	2661	14 AF172332	AF172332 Human her
35	133.5	4.0	10793	1 AE010748	AE010748 Methanosa
36	133.5	4.0	15274	1 AE004994	AE004994 Halobacte
37	133	3.9	1849	1 SGSTRBLF	X78973 S. galbus DS
38	132.5	3.9	10029	1 AE011825	AE011825 Xanthomon
39	132.5	3.9	11085	1 AE004450	AE004450 Pseudomon
40	132.5	3.9	25681	1 SABA14559	AJ14559 Saccharot
41	132.5	3.9	117096	9 AC008468	AC008468 Homo sapi
42	131.5	3.9	12730	1 HALPHSOD	M2454 H. halobium
43	131.5	3.9	12730	1 AE005055	AE005055 Halobacte
44	131.5	3.9	21391	1 AB070950	AB070950 Streptomy
45	131	3.9	13184	1 AE005096	AE005096 Halobacte

RESULT 1

ALIGNMENTS

ARI25096
 LOCUS ARI25096 2478 bp DNA linear PAT 16-MAY-2001
 DEFINITION Sequence 47 from patent US 6177075.
 ACCESSION ARI25096
 VERSION ARI25096.1 GI:14111158
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 2478)
 AUTHORS Christian,P.Daniel., Gordon,K.Hienrich,Julius. and
 TITLE Insect viruses and their uses in protecting plants
 JOURNAL Patent: US 6177075-A 47 23-JAN-2001;
 FEATUERS Location/Qualifiers
 source 1. 2478
 /organism="unknown"
 BASF COUNT 516 a 854 c 607 g 501 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 3.15e-211 Length: 2478
 Score: 3374.00 Matches: 647
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: Gaps: 0
 US-09-677-653a-50 (1-647) x ARI25096 (1-2478)

QY 1 MetGlyAspAlaGlyValAlaSerGlnArgProHisAsnArgGlyThrArgAsnVal 20
 Db ATGGGAATCTCTGGAGGGGCTACACAGCAGCTCACAAACGTCGCGAACCCGTAACGTT 425
 QY 21 ArgValSerAlaAsnThrValThrValAsnGlyArgArgAsnGlnArgArgThrGly 40
 Db CGGGTACGCCCAACACCCGTCACCGTCAATGCTAGAAAGAACACACGGCTCGGACCGGA 485
 QY 41 ArgGlnValSerProProAspAsnPheThrAlaAlaAlaGlnAspLeuAlaGlnSerLeu 60
 Db AGCCAGATTCTCCCTCGACCAATTTCACCGCTGCTGCACAAACGCTCGCGCAAGCCTT 545
 QY 61 AspAlaAsnThrValThrPheProAlaAsnIleSerSerMetProGluPheArgAsnThr 80
 Db GAGGCCAACACCGTCACTTCCCGGCTAACATCTAGCATGCGCGCAATTCCGGAATTGG 605
 QY 81 AlaIleGlyIleValIleAspLeuAspSerAspSerIleGlyTrpTrpPheIleValAsp 100
 Db GCCAAGGAAAGATCGACCTCGACCTCGATTCATCGGCTGGTACTTCAAGTACCTTGAC 665
 QY 101 ProAlaGlyAlaThrGlnSerAlaArgAlaValGlyGluTrpTrpSerIleProAspGly 120
 Db CCAAGCGGGTCTACAGAGTCTCGCGCGCGCTGGCGAGTACGAAAGATCCCTGACGGC 725
 QY 121 LeuValIlePheSerValAspAlaGluIleArgGluIleTrpAsnGluGluCysProVal 140
 Db CTCTGTAAGTCTCTCCGTCGACGAGATGAAGATCTATAACAGAGAGTCCCGCTC 785
 QY 141 ValThrAspValSerValProLeuAspGlyArgGlnTrpSerLeuSerIlePheSerPhe 160
 Db GTACACGACGTCCTCCCTCCCGTCGACGGCGCGCACTGGAGCTCTCGATTTCTCTTT 845
 QY 161 PrometPheArgThrAlaTrpValAlaValAlaAsnValGluAsnGlyGluMetSerLeu 180
 Db CCGATGTTCAAGAACCGCTACGTCGCGCGTACGCAACGTCGAGAAACAAGGATGTCGCTC 905
 QY 181 AspValValAsnAspLeuIleGluTrpLeuAsnAsnLeuAlaAspTrpArgTrpValVal 200
 Db GAGCTTGTCACGACCTCACTCGAGTGGCTCAACAACTCGCCGACGCGGTATGTCGTT 965
 QY 201 AspSerGluGlnTrpIleAsnPheThrAsnAspThrThrTyrrTrpValArgIleArgVal 220
 Db GACTCTGAACAGTGTAACTTACCAATGACACCACTACTACTGCGCATCCGCGTT 1025

QY 221 LeuArgProThrTrpAspValProAspProThrGlnGluLeuValArgThrValSerAsp 240
 Db CTACGTCACACCTACGACGATTCACAGACCCACAGAGGCGCTTGTCCACAGTCTCGAC 1085
 QY 241 TyrArgLeuThrTrpValAlaIleThrCysGluAlaAsnMetProThrLeuValAspGln 260
 Db TACCGCTCACTTATTAAGGGATGAATGAAACCAATGCAACATGCAACATCTGTCAGCAA 1145
 QY 261 GlyPheTrpIleGlyGlnTrpAlaLeuThrProThrSerLeuProGlnTrpAspVal 280
 Db GGGCTTGGATCGGGCGCAGTACGCTCACCCCGCTACCGCTACCGAGTACGAGCTC 1205
 QY 281 SerGlnAlaTrpAlaLeuHisThrLeuThrPheAlaArgProSerSerAlaAlaLeu 300
 Db AGCAGAGGCTACGCTCGACACACTTACCTCGCCAGACCATTCAGACGCGCTGCACCTC 1265
 QY 301 AlaPheValTrpAlaGlyLeuProGlnGlyGlyThrAlaProAlaGlyThrProAlaTrp 320
 Db GCGTTGTGTGGCAGGTTGGCCACAGGTTGGCAGCTCGCTGCAGACCTCCAGCCTGG 1325
 QY 321 GlnGlnAlaSerSerGlyGlyTrpLeuThrTrpArgHisAsnGlyThrThrPheProAla 340
 Db GAGCAGGATCTCGGGGTGCTACTCCTGCGCGCACACAGGTACTACTTCCACGCT 1385
 QY 341 GlySerValSerTrpValLeuProGluGlyPheAlaLeuGluIleArgTrpAspAsp 360
 Db GGCTCCCTTAGCTACGTTCTCCCTGAGAGTTGGCCCTTAGAGGCTACGACCCGAAAGAC 1445
 QY 361 GlySerTrpThrAspPheAlaSerAlaGlyAspThrValThrPheArgGlnValAlaVal 380
 Db GGGCTTTGGACCCACTTCCTCGCGAGACACGCTCACTTCCGGCGAGTCCGCGCTC 1505
 QY 381 AspGluValValValThrAsnAsnProAlaGlyGlySerAlaProThrPheThrVal 400
 Db GACGAGGCTCTGTGACCAACACCCCGGGGGGGGCGCCCGCCACTTACCGCTG 1565
 QY 401 ArgValProProSerAsnAlaTrpThrAsnThrValPheArgAsnThrLeuLeuGluThr 420
 Db AGAGTGCGCCCTTCAAAAGGCTTACACCAACACCGTGTATGGAACACGCTTAGAGACT 1625
 QY 421 ArgProSerSerArgArgLeuGluLeuProMetProProAlaAspPheGlyGlnThrVal 440
 Db CGACCTCTCTCTGTAGGCTCGAATCCCTATCCACCTGCTACTTGGACAGAGGCTC 1685
 QY 441 AlaAsnAsnProLysIleGlnGlnSerLeuLeuGluThrLeuGlyCysTrpLeuVal 460
 Db GCCAACACCCGAAAGATCGAGCTGCTTCTTAAAGAAACACTTGGCTGCTATTGGCTC 1745
 QY 461 HisSerLysMetArgAsnProValPheGlnLeuThrProAlaSerSerPheGlyAlaVal 480
 Db CACTCCAAATATGGAAACCCCGTTTCCAGCTCACGCGACGAGCTTGGCGCGCTT 1805
 QY 481 SerPheAsnAsnProGlyTrpGluArgThrArgAspLeuProAspTrpThrGlyIleArg 500
 Db TCTTTAAACATCTCGGATTATGAGCCACACGCGGACCTCCCGGATCACTGGCAATCCG 1865
 QY 501 AspSerPheAspGlnAsnMetSerThrAlaValAlaHisPheArgSerLeuSerHisSer 520
 Db GACTCATTTGACCAACATGTCACCGCTGTGCGCCACTTCCGCTCACTCTCCACTCC 1925
 QY 521 CysSerIleValThrLysThrTrpGlnGlyTrpGluGlyValAlaThrAsnValAsnThrPro 540
 Db TGCAGTATGTCATCTAAAGCTACAGGGTTGGGAAGGGGTACAGAACTCAACACAGCCT 1985
 QY 541 PheGlyGlnPheAlaHisAlaGlyLeuLeuLysAsnGluGluIleLeuCysLeuAlaAsp 560
 Db TTCGGCCAAATTCGCGACCGCGGCTCTCAAGAAATGAAGAAATCTCTGCTCGCGGAC 2045
 QY 561 AspLeuAlaThrArgLeuThrGlyValIleThrProAlaThrAspAsnPheAlaAlaVal 580
 Db GACTTGCCACCCGCTCTACAGGTTGTACCCCGCCACTGACAACTTCCGCGCGCTT 2105

QY 581 SerAlaPheAlaIalaSmetLeuSerSerValLeuLysSerGluAlaThrSerSerIle 600
|||||
Db 2106 TCGCCTTCGCCGGAACATGCTGCTCCGTGTAAGTCGAGGCAACGTCCTCCATC 2165
QY 601 IleLysSerValGlyLutThrAlaValAlaAlaGlnSerGlyLeuAlaLysLeuPro 620
|||||
Db 2166 ATCAAGTCGCTGGCGAGACTGCGGTGCGGCGGTCAAGTCGCGCTCGGGAAGCTACCC 2225
QY 621 GlyLeuLeuMetSerValProGlyLysIleAlaAlaArgValAlaArgAlaArgAlaArg 640
|||||
Db 2226 GGACTCTAATGATGCTACAGGGAAGATTCCGCGCTGTCGCGCGCGCGAGCGCGC 2285
QY 641 ArgArgAlaAlaArgAlaAsn 647
|||||
Db 2286 CGCCGCGCCCTCTGTCACAT 2306
RESULT 2
ARI25097 2478 bp DNA linear PAT 16-MAY-2001
LOCUS ARI25097 Sequence 49 from patent US 6177075.
DEFINITION ARI25097
ACCESSION ARI25097
VERSION ARI25097.1 GI:14111159
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2478)
AUTHORS Christian, P. Daniel., Gordon, K. Heinrich, Julius, and
Hanski, T. Nelson.
TITLE Insect viruses and their uses in protecting plants
JOURNAL Patent: US 6177075-A 49 23-JAN-2001;
FEATURES Location/Qualifiers
source 1..2478
BASE COUNT 516 a 854 c 607 g 501 t
ORIGIN
Alignment Scores:
Pred. No.: 3,15e-211 Length: 2478
Score: 3374.00 Matches: 647
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
US-09-677-653a-50 (1-647) x ARI25097 (1-2478)
QY 1 MetGlyAspAlaGlyValAlaSerGlnArgProHisAsnArgArgGlyThrArgAsnVal 20
|||||
Db 366 ATGGAGATGCTGGAGGGGCGACAGCGACTCACACCGTCGCGGAACCGCTAACGTT 425
QY 21 ArgValSerAlaAsnThrValThrValAsnGlyArgArgAsnGlnArgArgThrGly 40
|||||
Db 426 CGGGTAGGCGCCAAACCGCACCGTCAATGATGTAAGAAACCAACGCGCTCGACCGGA 485
QY 41 ArgGlnValSerProProAspAsnPheThrAlaAlaAlaGlnAspLeuAlaGlnSerLeu 60
|||||
Db 486 AGGCAAGTTTCTCCCTGCAATTTCACCGCTGCTGCAACAGACTTCGCGCAAGCCTT 545
QY 61 AspAlaAsnThrValThrPheProAlaAsnIleSerSerMetProGluPheArgAsnTrp 80
|||||
Db 546 GAGCGCAACACGCTACCTTCCCGGTAAACATCTAGCATGCGCGCAATTCCGGAATTGG 605
QY 81 AlaLysGlyLysIleAspLeuAspSerAspSerIleGlyTrpThrPheLysTrpLeuAsp 100
|||||
Db 606 GCCAAGGGAAGATCGACCTCGACTCGATTCATCGGCTGTAATTCACAGTACCTTGAC 665
QY 101 ProAlaGlyAlaThrGlnSerAlaArgAlaValAlaGlyLutThrSerLysIleProAspGly 120
|||||
Db 666 CCAAGCGGTCTCAAGATCTGCGCGCGCGCTGCGGAGACTCGAAGATCCCTGACGGC 725
QY 121 LeuValLysPheSerValAspAlaGluIleArgGluIleArgGluIleArgGluIleArg 140
|||||

Db 726 CTGTCAGATTCTCCGTCGACGAGATTAAGAGATCTATACGAGAGAGTCCCGCTC 785
QY 141 ValThrAspValSerValProLeuAspGlyArgGlnTrpSerLeuSerIlePheSerPhe 160
|||||
Db 786 GTCACTACAGTGTCCGTCCTCCCTCGACGGCGCCAGAGGAGCTCTCGATTTCTCCTTT 845
QY 161 PrometPheArgThrAlaTrpValAlaValAlaAsnValGlnLysLysGluMetSerLeu 180
|||||
Db 846 CCGATGTTCAAGACCGGCTACGTCGCGGTAGCAACGTCGAGAAAGAGATGTCCTC 905
QY 181 AspValValAsnAspLeuIleGluTrpLeuAsnLeuAlaAspTrpArgTrpValAla 200
|||||
Db 906 GAGCTTCTCAACGACCTCATCGAGTGGCTCAACAATCTCCGCACTGCTTATGTCCTT 965
QY 201 AspSerGluGlnTrpIleAsnPheThrAsnAspThrThrTrpTrpValArgIleArgVal 220
|||||
Db 966 GACTCTGAACAGTGAATTACTTACCAATGACACACGACTACGTCGCACTCCGCTT 1025
QY 221 LeuArgProThrTrpAspValProAspProThrGluGlyLeuValArgThrValSerAsp 240
|||||
Db 1026 CTACGTCCAACCTACGACGTTCCAGACCCACAGAGGCGCTTGTGCAACAGTCTCAGAC 1085
QY 241 TyrArgLeuThrTrpLysAlaIleThrCysGluAlaAsnMetProThrLeuValAspGln 260
|||||
Db 1086 TACCGCTCACTATATAAGGAGGATTAACATGTAAGCCACACTGCGACACTCGACCAA 1145
QY 261 GlyPheTrpIleGlyGlyGlnTrpAlaLeuThrProThrSerLeuProGlnTrpAspVal 280
|||||
Db 1146 GGCCTTTGGATCGCGCGCCGACGCTACCCGCTACCGCTACCGCGAGTACGACGCTC 1205
QY 281 SerGluAlaThrAlaLeuHisThrLeuThrPheAlaArgProSerSerAlaAlaLeu 300
|||||
Db 1206 AGCAGGCTACGCTGCGACACTTGTACTTGGCCAGACATCAAGCGCGGTGACCTC 1265
QY 301 AlaPheValTrpAlaGlyLeuProGlnGlyGlyThrAlaProAlaGlyThrProAlaTrp 320
|||||
Db 1266 GCGTTTGTCGAGGAGGTTGGCCACAGGAGGACACGCTGCGGACACTCCAGCCTGG 1325
QY 321 GluGlnAlaSerSerGlyGlyTrpLeuThrTrpArgHisAsnGlyThrThrPheProAla 340
|||||
Db 1326 GACGAGGCAATCCGCGGTGCTACCTACCTGCGCCGACCAACGTAATTTCCAGACT 1385
QY 341 GlySerValSerTrpValLeuProGluGlyPheAlaLeuGluLysArgTrpAspProAsnAsp 360
|||||
Db 1386 GGCCTCGCTTACGCTACGTTCCCTCGAGGGTTTGGCCCTTAGCGGCTACGACCCGAAGAC 1445
QY 361 GlySerTrpThrAspPheAlaSerAlaGlyAspThrValThrPheArgGlnValAlaVal 380
|||||
Db 1446 GGCCTTGAGACCGACTTCGCTCCGAGAGACACCGTCACTTCCGGAGGTCCGCGTC 1505
QY 381 AspGluValValValThrAsnAsnProAlaGlyGlyGlySerAlaProThrPheThrVal 400
|||||
Db 1506 GAGCAGGTCGTTGACCAACACCCCGCGGGGGGCGCCGCCACCTTACACCGTG 1565
QY 401 ArgValProProSerAsnAlaTrpThrAsnThrValPheArgAsnThrLeuLeuGluThr 420
|||||
Db 1566 AGAGTGCCCCCTTCAACCGCTTACACCAACACCGTGTATTGAGAACGCTTATAGAGACT 1625
QY 421 ArgProSerSerArgArgLeuGluLeuProMetProProAlaAspPheGlyGlnThrVal 440
|||||
Db 1626 CGAACCTCTCTGCTGAGGCTCGAATCCCTATGCAACCTGCTGATTTGGACAGAGGTC 1685
QY 441 AlaAsnAsnProLysIleGluGlnSerLeuLeuLysGluThrLeuGlyCysTrpLeuVal 460
|||||
Db 1686 GCCAACCAACCGAAGATCGAGCTGCTTCTTAAGAAACACTTGGCTCTATTGGTTC 1745
QY 461 HisSerLysMetArgAsnProValPheGlnLeuThrProAlaSerSerPheGlyAlaVal 480
|||||
Db 1746 CACTTCAAAATGCGAAACCCGTTTTCAGCTACGCGACACGACTCTTTGGCGCGCTT 1805
QY 481 SerPheAsnAsnProGlyTrpGluArgThrArgAspLeuProAspTrpThrGlyIleArg 500
|||||
Db 1806 TCTTTCAACAATCCGGGTTATGAGCGACACGCGGACTCCCGGACATACACTGCAATCCGT 1865

Qy	501	AASPSPRPheAspGlnAsnMetSerThrAlaValAlaHisPheArgSerLeuSerHisSer	520
Db	1866	GACTCATTTGCAGACCAACATGTCACCCGTCGTGGCCCACTTCGGCTCACTCCACATCC	1925
Qy	521	CysSerIleValI ThrIysThrTrpGlnGlyTTPGluGlyValIThrAsnValAsnThrPro	540
Db	1926	TGCAGATGCTACATAAGACTTACCAAGGTTGGAAAGGCTGACGAACCTCAACAGCCT	1985
Qy	541	PheGlyGlnPheAlaHisIleValLeuLeuLysAsnGluGluIleLeuCysLeuAlaAsp	560
Db	1986	TTCGGGCATTTCCGCGACCGCGGCGCTTCCTCAAGATATGAGAGATCTCTGCCCTCCGAC	2045
Qy	561	AspLeuAlaIThrArgLeuIThrGlyValITyPProAlaIThrAspAsnPheAlaIleVal	580
Db	2046	GACCTGGCCACCCGCTCTCAAGTGTCTTACCCCGCCACAGACAACTTCGCGGCCCTCCGCTT	2105
Qy	581	SerAlaPheAlaIleAsnMetLeuSerSerValLeuLysSerGluAlaIThrSerSerIle	600
Db	2106	TCTGCTTTCGCGGCAACATGCTGTCTCTCCGTCTGAACTCGGAGGCAACGTCTCCATC	2165
Qy	601	IleLysSerValIGlyCyluIThrAlaValGlyAlaIleGlnSerGlyLeuAlaLysLeuPro	620
Db	2166	ATCAAGTCCGTTGGGAGAGACTCCCTCGCGCGGCTCACTCGCCCTCGCGCAAGCTACCC	2225
Qy	621	GlyLeuLeuMetSerValIProGlyLysIleIleAlaIleArgValIThrAlaArgIThrArg	640
Db	2226	GGACTGCTATGATAGTGTACCAAGAAATGTCGCGGCGGTCTCGCGCGCGGACCGGC	2285
Qy	641	ArgArgAlaIleAlaArgAlaAsn	647
Db	2286	CGCCGGCGCGCTGTGCAT	2306
RESULT 3			
LOCUS	AR159968	2478 bp	DNA
DEFINITION	Sequence 1 from patent US 6251654.		linear
ACCESSION	AR159968		
VERSION	AR159968.1	GI:16222852	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	Unclassified.		
AUTHORS	1 (bases 1 to 2478)		
TITLE	Gordon,K.,Helinrich, and Hanzlik,T.,Neilson.		
JOURNAL	Modified small RNA viruses		
FEATURES	Patent: US 6251654-A 1 26-JUN-2001;		
	Location/Qualifiers		
	1..2478		
BASE COUNT	516 a	854 c	607 g
ORIGIN			501 t
	/organism="unknown"		
Alignment Scores:			
Score, No.:	3,15e-211	Length:	2478
Score:	3374.00	Matches:	647
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0
US-09-677-653A-50 (1-647) x AR159968 (1-2478)			
Qy	1	MetGlyAspAlaGlyValAlaISerGlnArgProHisAsnArgArgGlyIThrArgAsnVal	20
Db	366	ATGGGAGATGGCGGTGGCGTCACAGGCGACCTCAACACCTCGGGAACCCGTAAACGTT	425
Qy	21	ArgValISerAlaAsnIThrValIThrValAsnGlyArgArgAsnGlnArgArgIThrGly	40
Db	426	CGGGTCACCGCCACACCGCTCACCGTCAATGTGAAACAAACCAAGCGCGTCGGACCGGA	485
Qy	41	ArgGlnValSerProIAspAsnIThrAlaIleAlaGlnAspLeuAlaGlnSerLeu	60

D	486	AGCAAGTTTCTCCCTCCGACAAATTTACCGCGTGTGCACAGACACTCGCGCAAAAGCTT	545
Q	61	AspAlaAsnThrValThrPheProAlaAsnLLeSerSerMetProGluIbHeArgAsnTrp	80
D	546	GACGGCAACACCGTCACACTTTTCCCGGTAAACATCTTAGCATGGCCGAATTTCCGGAAATTGG	605
Q	81	AlaLysGlyLysLLeAspLeuAspSerAspSerLLeGlyYTrpYrPheLysTrileuAsp	100
D	606	GCCAAAGGGAAGATCGACTTCGACTCCGATTCATCGCGGTGACTCTCAAGATCCATTGAC	665
Q	101	ProAlaGlyIaThrGluSerAlaArgAlaValGlyLutYrSerLysLLeProAspGly	120
D	666	CCAGCGGGTGTACAGAGTCTGCCGGCGCGTGGGAGAGTACCGAAGATCCCTGACGGC	725
Q	121	LeuValLysPheSerValAspAlaGluIleArgGluIleYrAsnGluLutCySProVal	140
D	726	CTCGCAAGTTCTCCGTCGACGGCAGAGATTAAGAGATCATTAACAGAGAGATGCCCTTC	785
Q	141	ValTrpAspValSerValProLeuAspGlyArgGluTrpSerLeuSerLLePheSerPhe	160
D	786	GTCACTGACGTGCTCCGTCGCCCTTCGAGGGCGGCAATGGAGCCCTCGATTTTCTCTTT	845
Q	161	ProMetCPheArgThrAlaYrValAlaValAlaAsnValGluAsnLysGluMetSerLeu	180
D	846	CCGAGTTCAACACCGGCTACGTGCCGTAGCCACAGTGGAGACAGACAGATGTCCGCTC	905
Q	181	AspValValAsnAspLeuLLeGluTrpLeuAsnAsnLeuAlaAspTrpArgYrValVal	200
D	906	GACGTTGTCAACGACGACCTATCGATGGCTCAACAACTCGCGAGTGGGCTATATGCTGT	965
Q	201	AspSerGluGluTrpLLeAsnPheThrAsnAspThrThrYrYrValArgLLeArgVal	220
D	966	GACTGTGAACGTGGATTAACTTCACCAATGACACACGACTACGTCCGATTCGCGCTT	1025
Q	221	LeuArgProThrYrAspValProAspProThrGluGlyLeuValArgThrValSerAsp	240
D	1026	CTACGTCCAACCTACGACGTTCCAGACCCACAGAGGGCTTTGTGCAAGTCTCAAC	1085
Q	241	TyrArgLeuThrYrLysAlaLLeThrCysGluAlaAsnMetProThrLeuValAspGln	260
D	1086	TACGGCTCACTTATTAAGGCGATTAACATGTGAAGCCAAATGCAACTGTCGACCA	1145
Q	261	GlyPheTrpLLeGlyGlyGlnTrpAlaLeuThrProThrSerLeuProGluTrpAspVal	280
D	1146	GCGTTTGGAGCGGGCGGACGTACGCTCACCCGACATGACCTACCGGAGTACACATC	1205
Q	281	SerGluIaYrAlaLeuAsnLLeThrLeuThrPheAlaArgProSerSerLLeAlaLLeu	300
D	1206	AGCGAGGCTACGCTCTGCACTTTGACTTGGCCAGACATCCAGCGCGCTCATCTC	1265
Q	301	AlaPheValTrpAlaGlyLeuProGlnGlyGlyThrAlaProAlaGlyThrProAlaTrp	320
D	1286	GCGTTTGGTGGCAGAGTTTGGCCACAGGGGGCAGTGGCCGTGACGGCACTCACCTGG	1325
Q	321	GluGlnAlaSerSerGlyGlyYrLeuThrTrpArgHisAsnGlyYrThrPheProAla	340
D	1326	GAGCAGGATCTCGGGTGGCTACCTCACCCTGGCCCAACAGGTACTTCTCCCACT	1385
Q	341	GlySerValSerYrValLeuProGlnGlyPheAlaLeuGluArgYrAspProAsnSp	360
D	1386	GGCTCCGTTACTACGTTTCTCCCTGAGGGTTTCCCTTGACCGGTACGACCCGGAACAC	1445
Q	361	GlySerTrpThrAspPheAlaSerAlaGlyAspThrValThrPheArgGluValAlaVal	380
D	1446	GGCTCTTGGACCACTTCGCTTCCGAGAGACACCGTACTTCTCCGGCAGGTCCGCTC	1505
Q	381	AspGluValValValThrAsnAsnProAlaGlyGlyGlySerAlaProThrPheThrVal	400
D	1506	GAGCAGGCGTGTGGACCAACACCCGGGGGGGAGGAGGCCCCACCTCACCGTG	1565
Q	401	ArgValProProSerAsnAlaYrThrAsnThrValPheArgAsnThrLeuLeuGluThr	420
D	1566	AGAGTGGCCCCCTTCAACGCTTACACCAACACCGTGTATAGACACGCTCTTAGAGACT	1625

QY 421 ArpProserSerArgArgLeuLeuLeuProMetProProAlaAspPheGlyGlnThrVal 440
|||||
Db 1626 CGACCTCTCTCTGTAAGCTTGAACTCCCTATGCGACCTGCTGACTTGGACAGCGGCTC 1685
QY 441 AlaAsnAsnProLysIleGlnGlnSerLeuLeuLysGlnThrLeuGlyCysTyrLeuVal 460
|||||
Db 1686 GCCAACAACCCGAGAGTCGACACTCGCTTCTTAAAGAAACACTTGGCTGCTATTGGTTC 1745
QY 461 HisSerLysMetArgAsnProValPheGlnLeuThrProAlaSerSerPheGlyAlaVal 480
|||||
Db 1746 CACTCCAAATGCGAAACCCCGTTTCCAGCTCAGCCGACGACACTCTCTTGGCGCCCTT 1805
QY 481 SerPheAsnAsnProGlyTyrGlnArgThrArgAspLeuProAspTyrThrGlyTlleArg 500
|||||
Db 1806 TCCTTCACAACTCCGGGTTATGAGCGCACGCGGCTCCCGGACTACACTGGCATCCGT 1865
QY 501 AspSerPheAspGlnAsnMetSerThrAlaValAlaHisPheArgSerLeuSerHisSer 520
|||||
Db 1866 GACTCATTCGACGACAACATGTCACGCGCTGGCGCCACTTCCGCTCATCTCCACATCC 1925
QY 521 CysSerLleValIleThrLysThrTyrGlnGlyTyrPgluGlyValIleThrAsnValAsnThrPro 540
|||||
Db 1926 TGCAGTATCGTCACTAAGACCTACAGGCTTGGGAGGCGCTCAGCAAGCTCAACACGCT 1985
QY 541 PheGlyGlnPheAlaHisAlaGlyLeuLeuLysAsnGlnGluIleLeuCysAlaLeuAlaAsp 560
|||||
Db 1986 TTCGGCCAAATTCGCGACGCGGCGCTCTCTCAAGATGAGAGATCTCTGCTGCCGCGAC 2045
QY 561 AspLeuAlaThrArgLeuThrGlyValTyrProAlaThrAspAsnPheAlaAlaVal 580
|||||
Db 2046 GACTGGGCGACCGCTTCACAGGTGTCTACCCCGCACTGACAACTTGGCGCGCGCTT 2105
QY 581 SerAlaPheAlaAlaAsnMetLeuSerSerValLeuLysSerGluAlaThrSerSerLle 600
|||||
Db 2106 TCTGCTTCCGCGGCAACATCTGTCTCCCTGCTGAAGTGGAGGCAACGCTCTCCATC 2165
QY 601 IleLysSerValGlyGlnThrAlaValGlyAlaGlnIleSerGlyLeuAlaLysLeuPro 620
|||||
Db 2166 ATCAAGTCCGTTGGCGAGACTGCGGTGCGCGGCTCAGTCCGGCTCGCGAAGCTTACC 2225
QY 621 GlyLeuLeuMetSerValProGlyLysIleAlaAlaArgValArgAlaArgAlaArg 640
|||||
Db 2226 GGACTGGCTAATGAGTGTACAGGGAAGATTGCCGGGCTGTCCGGCGCGCGAGCGGCG 2285
QY 641 ArgArgAlaAlaArgAlaAsn 647
|||||
Db 2286 CGCCGCGCGCTCGTGCCTAAT 2306
RESULT 4
AX028950
LOCUS AX028950 2478 bp DNA linear PAT 16-SEP-2000
DEFINITION Sequence 1 from Patent WO974666.
ACCESSION AX028950
VERSION AX028950.1 GI:10189993
KEYWORDS
SOURCE Helicoverpa armigera stunt virus.
ORGANISM Helicoverpa armigera stunt virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Tetraviridae;
unclassified Tetraviridae.
1 (bases 1 to 2478)
REFERENCE
AUTHORS Gordon, K.H. and Hanzlik, T.N.
TITLE Modified small rna viruses
JOURNAL Patent: WO 974666-A 1 11-DEC-1997;
COMM SCIENT IND RES ORG (AU) ; GORDON KARL HEINRICH (AU) ; HANZLIK
TERRY NELSON (AU)
FEATURES
source location/qualifiers
1..2478
/organism="Helicoverpa armigera stunt virus"
/db_xref="taxon:37206"
BASE COUNT 516 a 854 c 607 g 501 t
ORIGIN

Alignment Scores:
Pred. NO.: 3,15e-211 Length: 2478
Score: 3374.00 Matches: 647
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 gaps: 0
US-09-677-653a-50 (1-647) x AX028950 (1-2478)
QY 1 MetGlyAspAlaGlyValAlaSerGlnArgProHisAsnArgArgGlyThrArgAsnVal 20
|||||
Db 366 ATGGAGATGCTGGAGTGGCGCTCACAGGACCTCAACAACGCTCGGGGAACCGTAAGCTT 425
QY 21 ArgValSerAlaAsnThrValThrValAsnGlyArgArgAsnGlnArgArgArgThrGly 40
|||||
Db 426 CGGGTCACGCGCAACACCGTCAACGTCATGTGTAGAAACCAACACGCGCTCGACCGGA 485
QY 41 ArgGlnValSerProProAspAsnPheThrAlaAlaGlnAspLeuAlaGlnSerLeu 60
|||||
Db 486 AGCGAAGTTTCTCCCTGACAAATTTCCACCGCTGCTGACAAAGACTTCGCGCAAGCTT 545
QY 61 AspAlaAsnThrValThrPheProAlaAsnIleSerSerMetProGluPheArgAsnThr 80
|||||
Db 546 GACGCGCAACACCGTCACTTTCGCGCTAACATCTCTACAGTCCGCAATTCGCGAATGG 605
QY 81 AlaLysGlyLysIleAspLeuAspSerAspSerIleGlyTyrTyrPheLysTyrLeuAsp 100
|||||
Db 606 GCCAAGGGAAGATCGACCTCGACTCCATTCATCGGCTGTGATTCMACTACTCTTGAC 665
QY 101 ProAlaGlyAlaThrGlnSerAlaArgAlaValGlyGluTyrSerLysIleProAspGly 120
|||||
Db 666 CGAGCGGTGCTACAGATCTGCGCGCGCGCTGCGGAGTACTGAAATATCTTGACGCG 725
QY 121 LeuValLysPheSerValAspAlaGluIleArgGluIleTyrAsnGlnGluCysProVal 140
|||||
Db 726 CTCGTCAGATTCTCCGCTCGACGACAGATTAAGAGATCTTAACGAGAGTGGCCCGCTC 785
QY 141 ValThrAspValSerValProLeuAspGlyArgGlnTyrPheSerLysIlePheSerPhe 160
|||||
Db 786 GTCACTGACGTGCTCGTCCCTCCCTCGACGCGCGCAAGTGGACCTCTCATTTTCTCTT 845
QY 161 PrometPheArgThrAlaTyrValAlaValAlaAsnValGluAsnLysGlnMetSerLeu 180
|||||
Db 846 CCGATGTCACAGACCGCTACGTGCGCTGACGAGACGTGAGAACAGAGATGTCGCTC 905
QY 181 AspValValaAsnAspLeuIleGluThrPheAsnAsnLeuAlaAspTyrArgTyrValVal 200
|||||
Db 906 GACGTGTCAACGACCTCATGTGAGTGGCTCAACAATCTCGCGACTGGCGTTATGTGCTT 965
QY 201 AspSerGlnGlnTyrIleAsnPheThrAsnAspThrThrTyrTyrValArgIleArgVal 220
|||||
Db 966 GACTCTGAACAGTGTGATTAACCTTCAACAATACACACAGTACTGTCCGATCCGCGCTT 1025
QY 221 LeuArgProThrTyrAspValProAspProThrGlnGlyLeuValArgThrValSerAsp 240
|||||
Db 1026 CTACGTCACCTTACGAGCTTCCACAGCCCAAGAGGCGCTTGTTCGACAGTCTCAAC 1085
QY 241 TyrArgLeuThrTyrLysAlaIleThrCysGluAlaAsnMetProThrLeuValAspGln 260
|||||
Db 1086 TACCGCTCACTTAAAGGCGATTAACATGTGAAGCAACATGCCAACACTGTGACCA 1145
QY 261 GlyPheThrIleGlyGlyGlnTyrAlaLeuThrProThrSerLeuProGlnTyrAspVal 280
|||||
Db 1146 GCGTTTGTGATCGGGCGGCGCACTACGCTCACCCGACTACCTAACCCAGTACGACGTC 1205
QY 281 SerGluValTyrAlaLeuHisThrLeuThrPheAlaArgProSerSerAlaAlaAlaLeu 300
|||||
Db 1206 AGCGAGGCTACGCTGTGCAACCTTGTGACCTTGCACACCATTCACAGCGCGGTACATC 1265
QY 301 AlaPheValIleThrAlaGlyLeuProGlnGlyGlyThrAlaProAlaGlyThrProAlaThr 320
|||||
Db 1266 GCGTTTGTGTGGCAGGTTTGCACAGGTTGGACACTGCGCTGCAAGGACTTCAGCCTGG 1325

```

OY 321 GUGLINALASERSEGLYTYRLEUTHRTPARGHISANGLYTHRPHPROALA 340
    |||||||
DB 1326 GAGACAGCATCTCGGTGGCTACTCTACCTGGCCACAAACGCTACTTCCCGACT 1385
OY 341 GYSEVALSERVYRVALLEUPROGLUGLYPHEALALEUGLUARGTYRASPPOASNAP 360
DB 1386 GGCTCCGTTAGTACGTTCTCCCTGAGAGGGTTTGCCCTTGAGGCGTACGACCCGAAGGAC 1445
OY 361 GLYSERTTPHRSAPHEALASERIALAGLYASPTHYVALTHRPHARGGLINVALALVAL 380
    |||||||
DB 1446 GGCTCTGGACCGACTTCGCTCCGACGAGACCGCTCACTTCCGGCAGTCCGCGCTC 1505
OY 381 ASPGIYVALVALYVALTHRASNAPPROALAGLYGLYSERIALAPROTHRPHETHVAL 400
    |||||||
DB 1506 GACGAGTGTCTTGACCAACCAACCCCGCGGGCGGACGCCGCCCTTACCCTG 1565
OY 401 ARGVALPROBROSERASNALATYRTHRASNTHVALPHEARGANTHRLEULEUGLUTHR 420
    |||||||
DB 1566 AGAGTCCCCCTTCAAAACGCTTACACCAACACCGTGTAGGAACAGCTTAGAGACT 1625
OY 421 ARGPROSERSEARGARGLEUGLULEUPROMETPROPROALASPHLEGUINTHVAL 440
    |||||||
DB 1626 CGACCTCTCTCTGAGGCTCGAAGTCCCTATGCCACTGCTTGGACAGAGCGTCC 1685
OY 441 ALAASNAPPROLYSILEGLINSERLLEULEUGLUTHRLEUGLYCYSTYRLEUVAL 460
    |||||||
DB 1686 GCCAACACCCGAGATCGAGCGACTGCTTTAAAGAAACACTTGGCTCTATTGGTCC 1745
OY 461 HISSELYSEMEARGANPROVALPHEGLNLEUTHRPROALASERSEPHGLYALVAL 480
    |||||||
DB 1746 CACTTCAAAATGCGAAACCCCTTTTCCACTCCACCGCACCGCTTGGCGCGCTT 1805
OY 481 SECPHASNAPPROGLYTYRGLUARGTYRARGASPLEUPROSPYRTHRGUYLEARG 500
    |||||||
DB 1806 TCCTTAACAATCCGGGTTATGAGCGACACGGGACCTCCGGGACTACATGCGATCCGT 1865
OY 501 ASPSERPHEASPGIINASNMESETRHVALALVALALHISPHEARSELEUSERHISSE 520
    |||||||
DB 1866 GACTCATTTGGACCAAGATGTCACCGGCTGCGGCCACTTCCGCTCCTCCACTCC 1925
OY 521 CYSSERTLEVALTHRYSTHRYTYRGLINGLYTTPGLUGLYVALTHRASNVALASNTHPRO 540
    |||||||
DB 1926 TGCAGTATGCTCACTAAGACTCTACAGGGTGGGAAGGCTCTCAAGAACCTCAACGCGCT 1985
OY 541 PHEGLYGLNPHLEALHISLAGLYLEULEUASNGLULEUCYSLLEUVALAASP 560
    |||||||
DB 1986 TTGGCCAAATCGGCAACGGGGCTCTCAAGATGAGAGATCCCTGCTCCGCGGAC 2045
OY 561 ASPLEUALATHRARGLEUTHRGLYVALTYRPROALATHRASNAPNHEALALALVAL 580
    |||||||
DB 2046 GACTGGCCACCGGCTCAAGAGTGTCTACCCGCCACTGACAACTTCCGGCGCGCTT 2105
OY 581 SERIALPHEALALASNMETLEUSERSERVALLEULYSERGLUALATHRSERTILE 600
    |||||||
DB 2106 TCTGCTTGGCCCGAAGCTGCTCTCTGCTCAAGTCCGAGGCAACGCTCTCATC 2165
OY 601 ILELYSERVALGLYUPTHRALVALAGLYALAGLINSERGLYLEUVALYSLLEUPRO 620
    |||||||
DB 2166 ATCAATCTCGGTTGGCAGACTGCGTGGGGCGGCTCAAGTCCGCGCGGGAAGCTACCC 2225
OY 621 GLYLEULEUMETSERVALPROGLYSLLEALALARGVALARGLAARGALALARG 640
    |||||||
DB 2226 GGACTCTGATAGTGTACAGGGAAGATTCCGCGCGGTGTCCGCGCGGCGAGGGCGC 2285
OY 641 ARGATGALALALARGALASN 647
    |||||||
DB 2286 CGCGCGCGCGCTGTGCAAT 2306

```

```

RESULT 5
AX028971/c AX028971 2478 bp DNA linear PART 16-SEP-2000
LOCUS DEFINITION Sequence 22 from Patent WO9746666.

```

```

ACCESSION AX028971 GI:10190009
VERSION AX028971.1
KEYWORDS
SOURCE Helicoverpa armigera stunt virus.
ORGANISM Helicoverpa armigera stunt virus.
REFERENCE 1 (bases 1 to 2478)
AUTHORS Gordon, K.H. and Hanzlik, T.N.
TITLE Modified small rna viruses
JOURNAL Patent: WO 9746666-A 22 11-DEC-1997;
TERRY NELSON (AU)
FEATURES
    source location/Qualifiers
        1..2478 /organism="Helicoverpa armigera stunt virus"
        /db_xref="taxon:37206"
BASE COUNT 501 a 607 c 854 g 516 t
ORIGIN
Alignment Scores:
Pred. No.: 3,15e-211 Length: 2478
Score: 3374.00 Matches: 647
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-09-677-653A-50 (1-647) x AX028971 (1-2478)
OY 1 MetGLYASPLAGLYVALALASERGINARGPROHISASNARGGLYTHRARGSNVAL 20
    |||||||
DB 2113 ATGGAGATGCTGAGAGGGGCTCAGACGACCTCACAAACGCTGGGGAACCCGTAACGTT 2054
OY 21 ARGVALSERIALASNTHRVALTHRVALASNGLYARGARGASNGLNARGATGATGTYRGLY 40
    |||||||
DB 2053 CGGGTAGCGCCCAACACCGCTCACCGTCAAGTAGAAGAAACCAACGGCGTCCGACCGGA 1994
OY 41 ARGGLNVALSERPROPOASNPHETHRALALALAGLINSPLAULAGLINSERLEU 60
    |||||||
DB 1993 AGGCAAGTTTCTCCCTTGACATTTTCACCGCTGCTGACCAAGACCTCGCGCAACCTT 1934
OY 61 ASPALASNTHRVALTHRPHPROALASNILESESERMEPRROGLUPHARGASANTP 80
    |||||||
DB 1933 GACGCCAACACCGTCACTTCCCGCTACATCTACATGTCGCGCAATTCGGAATTGG 1874
OY 81 ALALYGLYLYSLLEASPLEUASPSERSETILEGYTTPYRTHRYSTYRLEUASP 100
    |||||||
DB 1873 GCCAAGGGAAGATCGACCTCGACTCGATTCCATGGCTGGTCAAGTACCTTGAC 1814
OY 101 PROALAGLYLATHRGLUSERIALARGALALVALAGLYLTYRSERTYSLLEPROASPGLY 120
    |||||||
DB 1813 CCAGCGGGTCTACAGAGTCTGCGCGCGCGCTGGCGAGTACTCGAAGATCCCTGACGGC 1754
OY 121 LEUVALYSPHESERVALASPLAGLUILEARGLUILETYRASNGLUCYSPROVAL 140
    |||||||
DB 1753 CTGCTCAAGTTCTCCGCGACGCGAGATPAAGAGATCATPAACGAGAGTCCCGCTC 1694
OY 141 VALTHRASPYALSERVALPROLEUASPGLYARGGLINTPRTSERLEUSERILEPHESE 160
    |||||||
DB 1693 GTTCACTGAGCTGCTCCCTCCGACGGCGCGCAATGGAGCTTCGATTTCCTCTTT 1634
OY 161 PROMETPHEARGTHRALATYRVALALALASNVALAGLUNSNLYSGLUSETSERLEU 180
    |||||||
DB 1633 CGGATGTTCAAGAACCGGCTTACGCTCGCGCTAGCAACGCTGAGAGAACAGAGATGTGCTC 1574
OY 181 ASPVALVALASNAPLEULIEGLUTRPLEUASNPLAULASPTTRPARGTYRVALVAL 200
    |||||||
DB 1573 GACGTTGTCAAGCACTCATCGAGTGGCTCAACAATCTGCGCGACGAGGATTATGTCGTT 1514
OY 201 ASPSERGLUGLINTRPLEASNPHETHRASNAPTHRTHRYTYRVALARGILEARGVAL 220
    |||||||
DB 1513 GACTCTGAACAGATGATTAATTACCAATGACACACGACTACTACGTCGCGATCCCGCTT 1454

```

QY 221 LeuArgProThrTyrAspValProAspProThrGluGlyLeuValArgThrValSerAsp 240
 DB 1453 CTACGCTCAACCTACGACGCTCCAGACCCACAGAGGGCTTGTCCACAGCTCTCAGAC 1394
 QY 241 TyrArgLeuThrTyrTysAlaIleThrCysGluAlaAsnMetProThrLeuValAspGln 260
 DB 1393 TACCGCTCACTATAAGGGGATACATGTGAAGCCAAACATGGCAACACTCGTCGACCAA 1334
 QY 261 GlyPheTrpIleGlyGlyGlnTyrAlaLeuThrProThrSerLeuProGlnTyrAspVal 280
 DB 1333 GGGTTTGGATCGCGCGCCGAGCTACGCCCGACTGACCTACCGCAGTGCAGAGTGC 1274
 QY 281 SerGlnAlaTyrAlaLeuHisThrLeuThrPheAlaArgProSerSerAlaAlaLeu 300
 DB 1273 AGCGAGGCTACGCTCTGCACTGTGACTTGGCTGCGACACTTCACAGCGCGCTGCACTC 1214
 QY 301 AlaPheValTrpAlaGlyLeuProGlnGlyGlyThrAlaProAlaGlyThrProAlaTrp 320
 DB 1213 GCGTTTGTGGGAGGTTTGGCCACAGGGTGGCACTGCGCTGAGGCACTCCAGCCTGG 1154
 QY 321 GluGlnAlaSerSerGlyGlyTyrLeuThrTrpArgHisAsnGlyThrThrPheProAla 340
 DB 1153 GACGAGGCACTCGGGGTGGCTACTCCTGAGCGCCACAAAGGTACTTCCAGCT 1094
 QY 341 GlySerValSerTyrValLeuProGlnGlyPheAlaLeuGlnArgTyrAspProAsp 360
 DB 1093 GGGTCGGTAAAGTACGCTTCCCTGAGAGGTTTGGCCCTTGAAGCGTACGACCGGAAGAC 1034
 QY 361 GlySerTrpThrAspPheAlaSerAlaGlyAspThrValThrPheArgGlnAlaVal 380
 DB 1033 GGGTCTTGGAGCAGCTTCCGTCGCGAGAGACACCTCACTTCCGGCAGTCCGCGTC 974
 QY 381 AspGluValValThrAsnAsnProAlaGlyGlySerAlaProThrPheThrVal 400
 DB 973 GACGAGGCTGTGACCAACAACCGCGCGGCGGAGCGCCCACTTCCAGCGTG 914
 QY 401 ArgValProProSerAsnAlaTyrThrAsnThrValPheArgAsnThrLeuGlnThr 420
 DB 913 AGAGTGGCCCTTCAACGGTTTACCAACACCGTGTGTGAAGAACAGCGCTTAGAGACT 854
 QY 421 ArgProSerSerArgArgLeuGluLeuProMetProProAlaAspPheGlyGlnThrVal 440
 DB 853 CGACCCCTCTCTGTAGGCTCGAACTCCATGCACTCTGACTTGGACACAGCGTGC 794
 QY 441 AlaAsnAsnProLysIleGluGlnSerLeuLeuLysGlnThrLeuGlyCysTyrLeuVal 460
 DB 793 GCCAACAAACCGAAGATCGAGAGTGGCTTCTTAAAGAAACACTTGGCTCTATTGGTC 734
 QY 461 HisSerLysMetArgAsnProValPheGlnLeuThrProAlaSerSerPheGlyAlaVal 480
 DB 733 CACTCCAAATGGGAAACCCCGTTTTCACACTCAGCGCACGCTCTTGGCGCGT 674
 QY 481 SerPheAsnAsnProGlyTyrGluArgThrArgAspLeuProAspTyrThrGlyIleArg 500
 DB 673 TCCCTTCAACAATCGGGTTATGAGCGCACAGCGACTCCCGGACTACACTGGCATCCG 614
 QY 501 AspSerPheAspGlnAsnMetSerThrAlaValAlaHisPheArgSerLeuSerHisSer 520
 DB 613 GACTCATATGACCAAGACATGTCACCGGTGGGCCCACTTCCGCTCACTCCCACTCC 554
 QY 521 CysSerTleValThrLysThrTyrGlnGlyTyrGluGlyValThrAsnValAsnThrPro 540
 DB 553 TGGAGTATCTCACTAAGACTCAAGAGGTGGGAAAGCGCTCAAGAACGCAACACGCT 494
 QY 541 PheGlyGlnPheAlaHisAlaGlyLeuLeuLysAsnGlnGluIleLeuCysLeuAlaAsp 560
 DB 493 TTGGGCAATTCGGCGACGGGGCTCTCTCAAGATAGAGATCTCTGCTCGCGGAC 434
 QY 561 AspLeuAlaThrArgLeuThrGlyValTyrProAlaThrAspAsnPheAlaAlaVal 580
 DB 433 GACCTGGCCACCGGTCTACAGAGTGTCTACCCGCGCACTGACAACTTCCGCGCGCGT 374

QY 581 SerAlaPheAlaAlaAsnMetLeuSerSerValLeuLysSerGluAlaThrSerSerIle 600
 DB 373 TCTGCTTCCCGCGAACAATGCTCTCTCCGCTGTAAGTCCGAGCAAGCTCTCCATC 314
 QY 601 IleLysSerValGlyGlnThrAlaValAlaAlaAlaGlnSerGlyLeuAlaLysLeuPro 620
 DB 313 ATCAAGTCCGTGGCGAGACTGCGGTGGCGGCTCAGTCCGAGCTCGGGAAGCTACCC 254
 QY 621 GlyLeuLeuMetSerValProGlyLysIleAlaAlaLarValArgAlaArgAlaArg 640
 DB 253 GGACTGCTAATGAGTGTACAGGGAAGATTGCGCGCTGTCCGCGCGCGAGCGCGC 194
 QY 641 ArgArgAlaAlaArgAlaAsn 647
 DB 193 CGCCGCGCGCTCTGTCCAAT 173
 RESULT 6
 HVScoat
 LOCUS 2478 bp ss-RNA linear VRL 11-MAY-1995
 DEFINITION Helicoverpa armigera stunt virus RNA2 coat protein (p71) and p17
 ACCESSION gene, complete cds.
 VERSION L37299.1 GI:609503
 KEYWORDS coat protein; transfer RNA-like structure.
 SOURCE Helicoverpa armigera stunt virus (individual isolate black mountain) cDNA to genomic RNA.
 ORGANISM Helicoverpa armigera stunt virus
 Viruses; ssRNA positive-strand viruses, no DNA stage; Tetraviridae; unclassified Tetraviridae.
 1 (bases 1 to 2478)
 REFERENCE Hanzlik, T.N., Dorrian, S.J., Johnson, K.N., Brooks, E.M. and Gordon, K.H.
 TITLE Sequence of RNA2 of the Helicoverpa armigera stunt virus (Tetraviridae) and bacterial expression of its genes
 JOURNAL J. Gen. Virol. 76 (Pt 4), 799-811 (1995)
 MEDLINE 97201539
 PubMed 9049325
 FEATURES
 source location/Qualifiers
 1..2478 /organism="Helicoverpa armigera stunt virus"
 /isolate="black mountain"
 /db_xref="taxon:37206"
 /clone="hasvrna2"
 1..282
 1..50
 5'UTR
 stem_loop
 /note="stem loop which has hexamer sequence, gguaaa, in same place on a stem loop structure on RNA1; putative"
 /function="unknown"
 283..756
 /gene="p17"
 283..756
 /gene="p17"
 /note="has PEST characteristics, makes tube-like structures when expressed in bacteria, unstable when expressed in baculovirus, start codon in poor context so is poorly expressed if at all; putative"
 /codon_start=1
 /protein_id="AAC37884.1"
 /db_xref="GI:609504"
 /translation="MSERTIAHSITTPPYTIALIPPEDEAGEMLEMRHSULTYVAE
 PTFGSAPIPSBWEETINGVGEKFLDTISPLHKSRKALTPSLSTSLAC
 PMSGIGRERSSTIPSAGSTSTLQRIQVLOSIRASASTRSLTASSSSPSPTOR"
 366..2309
 /gene="p71"
 366..2309
 /gene="p71"
 /standard_name="HasV coat protein"
 /note="coat protein for HasV, forms T=4 capsids with 240 copies of protein per capsid, upon assembly into particle is cleaved at N/F at residue positions 575-76"
 /codon_start=1
 /evidence="experimental"

DB	Accession	Gene	Protein	Length	Score	Percent Similarity	Best Local Similarity	Query Match
DB	141	ValThrAspValSerValProLeuAspGlyArgGlnTrpSerLeuSerIlePheSerPhe	160	786	786	100.00%	100.00%	100.00%
DB	121	LeuValLysPheSerValAspPalaGluIleArgGluIleTyrAsnGluGluCysProVal	140	726	726	100.00%	100.00%	100.00%
DB	606	GCCAAAGGAAAGATCGACCTCGATCCGATTCCTTCGGCTGGTCACTTCACCTTACCTTGAC	665	606	606	100.00%	100.00%	100.00%
DB	426	CGGGTCACGCGCACACCGCTCACCGTCAATGTAAGAAACCAACGCGCTCGGACCGGA	485	426	426	100.00%	100.00%	100.00%
DB	41	ArgGlnValSerProProAspAsnPheThrAlaAlaGlnAspLeuAlaGlnSerLeu	60	486	486	100.00%	100.00%	100.00%
DB	61	AspPalaAsnTrpValThrPheProAlaAsnIleSerSerMetProGluPheArgAsnTrp	80	546	546	100.00%	100.00%	100.00%
DB	81	AlaLysGlyLysIleAspLeuAspSerAspSerIleGlyTrpTyrPheLysTyrLeuAsp	100	606	606	100.00%	100.00%	100.00%
DB	101	ProAlaGlyLysIleAspLeuAspSerAspSerIleGlyTrpTyrPheLysTyrLeuAsp	120	666	666	100.00%	100.00%	100.00%
DB	121	LeuValLysPheSerValAspPalaGluIleArgGluIleTyrAsnGluGluCysProVal	140	726	726	100.00%	100.00%	100.00%
DB	141	ValThrAspValSerValProLeuAspGlyArgGlnTrpSerLeuSerIlePheSerPhe	160	786	786	100.00%	100.00%	100.00%

QY	161	ProMetPheAqgThrAlaTYrValAlaValAlaAsnValGluAsnLysGluMetSerLeu	180
Db	846	CGAGTTCACAAACCGCCCTACGTCGCCGTAGCACAAGTCGAGAACAGAGATGTCGCTC	905
QY	181	AspValValAlaAsnAPLeuLLeuLTPLeuAsnAsnLeuAlaSprPrpArgTYrVal	200
Db	906	GACGTGTGCACAGACTATCGATGAGTGGCTACAAACATCGCCAGCTGGGGTTATGTCGT	965
QY	201	AspSerGluGlnTPLeuAsnPheThrAsnAspThrTYrTYrValArgLLeaArgVal	220
Db	966	GACTCTGAACAGTGGATTAACTTCACCAATGACACACAGTACGTACGTCGCGGTT	1025
QY	221	LeuArgProThrTYrAspValProAspProThrGluGlyLeuValArgTYrValSerAsp	240
Db	1026	CTACCTCCAACTCAGACGAGTTCAGACCCACAGAGGGCTGTGTGCACACTCTCAAC	1085
QY	241	TYrArgLeuThrTYrLysAlaLLeThrCysGluAlaAsnMetProThrLeuValAspGln	260
Db	1086	TACGGCTCACTATTAAAGGCGATACATGTGAAGCCAACTGCAACACTCTGTCAACCA	1145
QY	261	GlyPheTrpLLeuGlyGlyGlnTYrAlaLeuThrProThrSerLeuProGlnTYrAspVal	280
Db	1146	GGCTTTGGATCGGGCGGCAAGTACGCTCACCAGCAGTACGCTACCGAGTACAGCTC	1205
QY	281	SerGluAlaTYrAlaLeuHisThrLeuThrPheAlaArgProSerSerAlaAlaLeu	300
Db	1206	AGCGAGGCTCAGCTCAGCACTTTCACCTTCGCGACACATCCAGCCCGCTCAGCTC	1265
QY	301	AlaPheValTrpAlaGlyLeuProGlnGlyTYrThAlaProAlaGlyThrProAlaTrp	320
Db	1266	GCGTTTGTGTGGGCAAGTTTCCACAGAGTGGCACTGCGCTGCAAGCACTCAGCTCG	1325
QY	321	GluGlnAlaSerSerGlyGlyTYrLeuThrTrpArgHisAsnGlyThrThrPheProAla	340
Db	1326	GAGCAGGATCCTCGGGGTACCTCAGCTCAGCGCCACAAAGGTACTCTCCAGCT	1385
QY	341	GlySerValSerTYrValLeuProGluGlyPheAlaLeuGluAlaArgTYrAspProAsnAsp	360
Db	1386	GGCTCCGTTACTACGTTCCTCCGTGAGGGTTTGGCCCTTAGAGCGTACACCCGCAAGAC	1445
QY	361	GlySerTrpThrAspPheAlaSerAlaGlyAspThrValThrPheArgGlnValAlaVal	380
Db	1446	GGCTCTTGGACCGACTTCGCTCCGACGAGACACCGTACTTCCGCGAGGTCCGCTC	1505
QY	381	AspGluValValThrAsnAsnProAlaGlyGlySerAlaProThrPheThrVal	400
Db	1506	GACGAGGTCTGTGGACCAACACCCCGCGGGGGGAGGGCCCCACTTCACCGTG	1565
QY	401	ArgValProProSerAsnAlaTYrThrAsnThrValPheArgAsnThrLeuLeuGluThr	420
Db	1566	AGAGTGGCCCCCTTCAAAACGCTTACACCAACCCGTGTTAGAACACGCTTAGAGACT	1625
QY	421	ArgProSerSerArgArgLeuGluLeuProMetProProAlaAspPheGlyGlnThrVal	440
Db	1626	CGACCTCTCTCTCTAGGGTCGAAACCTCATGCCACCTGCTGACTTTGGACAGACGGTC	1685
QY	441	AlaAsnAsnProLysLLeuGlnSerLeuLeuLysGluThrLeuGlyCysTYrLeuVal	460
Db	1686	GCCAACCAACCGAAGATCGAGCGAGTCTTTAAAGAAACCTGGCTGCTATTGGTCTC	1745
QY	461	HisSerLysMetArgAsnProValPheGlnLeuThrProAlaSerSerPheGlyAlaVal	480
Db	1746	CAGTCCAAATGGGAACCCCGTTTTCAGAGCTCAGCCAGCAGCTCTTTGGGGCGGTT	1805
QY	481	SerPheAsnAsnProGlyTYrGluArgThrArgAspLeuProAspTYrThGlyTYrLea	500
Db	1806	TCTCTTCAACATCCGGGTATTGAGCCACACCGGCACTCCCGGATACACTGGATCCGT	1865
QY	501	AspSerPheAspGlnAsnMetSerThrAlaValAlaHisPheArgSerLeuSerHisSer	520
Db	1866	GACTCATTTGACCAAGAACATGTCACCGGTGGGCCCACTTCCGCTCACTCCACATCC	1925

QY 521 CysSerIleValThrLysThrTyrgInGIYTrpGIuGIYValThrAsnValAsnThrPro 540
|||||
Db 1926 TGCAGTATCTGCTACTAAGACCTACAGGTTGGGAAGCGCTCACCAAGCTCAACACGCTT 1985
QY 541 PheGlyInPheAlaHisAlaGlyLeuLeuLysAsnGIuGIuIleLeuCysLeuAlaAsp 560
|||||
Db 1986 TTCGGCAATTCGGGCACAGGGGGCTCCTCAAGAAATGAGGATCGTCTGCTGGCGGAC 2045
QY 561 AsPheAlaThrArgLeuThrGlyValTyProAlaThrAspAsnPhelaAlaAlaVal 580
|||||
Db 2046 GACTGGCCACCGCTCTCAAGAGTGTCTACCCCGCCACTGACACTTCGCGGGCGCGT 2105
QY 581 SerAlaPheAlaAlaAsnMetLeuSerSerValLeuLysSerGIuAlaThrSerSerIle 600
|||||
Db 2106 TCTGCTTCGCGCGCAACAGTCTCTCTCGTCTGAAGTGGAGGACCTCTCCATC 2165
QY 601 IleLysSerValGIuGIuThrAlaValAlaGIYAlaAlaInSerGIYLeuAlaLysLeuPro 620
|||||
Db 2166 ATCAAGTCCGTGGCGAGACTGCGCGTGGCGGCTCAGTCCGGCTCGCAACCTACCC 2225
QY 621 GlyLeuLeuMetSerValProGIYLysIleAlaAlaArgValAlaArgAlaArg 640
|||||
Db 2226 GGACTGTAAATGAGTGTACCAAGGAAAGATTCGCGCGTGTCCGCGCGCGGCGGCGC 2285
QY 641 ArgArgAlaAlaArgAlaAsn 647
|||||
Db 2286 CGCGCGCGCTCGTCCCAAT 2306

RESULT 7
ARI25098 2479 bp DNA linear PAT 16-MAY-2001
LOCUS ARI25098
DEFINITION Sequence 51 from patent US 6177075.
ACCESSION ARI25098
VERSION ARI25098.1 GI:14111160
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2479)
AUTHORS Christian,P.Daniel., Gordon,K.Henrich,Julius. and
Hanzlik,T.Nelson.
TITLE Insect viruses and their uses in protecting plants
JOURNAL Patent: US 6177075-A 51 23-JAN-2001;
FEATURES
source location/Qualifiers
1..2479
BASE COUNT 516 a 855 c 607 g 501 t
ORIGIN

Alignment Scores:
Pred. No.: 2.23e-210 Length: 2479
Score: 3361.00 Matches: 647
Percent Similarity: 99.85% Conservative: 0
Best Local Similarity: 99.85% Mismatches: 0
Query Match: 99.61% Indels: 1
DB: Gaps: 0

US-09-677-653a-50 (1-647) x ARI25098 (1-2479)
QY 1 MetGlyAspAlaGlyValAlaSerGIuArgProHisAsnArgArgGIYThrArgAsnVal 20
|||||
Db 366 ATGGGAATGCTGAGAGGCGCTCACAGCACTACAAACCTCGCGGAGACCCGTAACGTT 425
QY 21 ArgValSerAlaAsnThrValThrValAsnGIYArgArgAsnGIuArgArgArgThrGly 40
|||||
Db 426 CGGCTCAGCGCCAAACACCGTCAATGTGTAGAGAAACCAAGCGCGTGGAGCCGA 485
QY 41 ArgGIuValSerProAspAsnPhelThrAlaAlaInSerGIYLeuAlaGlnSerLeu 60
|||||
Db 486 AGGGAAGTTCCTCCCTGCAATTTCAACGCTGTGCACAAAGACCTCGCCAAAGCCTT 545
QY 61 AspAlaAsnThrValThrPhe-ProAlaAsnIleSerSerMetProGIuPheArgAsnTr 80
|||||

Db 546 GAGCCAAACACCGTCACTTCCCGGTAACATCTTAGATGCGCGAATTCGGAATG 605
QY 80 PalaGlyLysIleAspLeuAspSerSerIleGlyTrpTyrPheLysTyLeuAs 100
|||||
Db 606 GGCCAAAGGAAGATTCAGACCTCGACTCGATTCCATCGGGCTGACTTCAAGTACCTTGA 665
QY 100 PProAlaGIYAlaThrGlnSerAlaArgAlaValGIYGIuThrSerLysIleProAspGI 120
|||||
Db 666 CCCAGCGGTGTCAACAAGTGTGGCGCGCTGTGGGAGAGTCAAGATCCCTGACGG 725
QY 120 YLeuValLysPheSerValAspAlaGIuIleArgGIuIleTyrsnGIuGIuCysProVa 140
|||||
Db 726 CTTGCTCAAGTCTCCCTCGACGCAAGATAGAGACTATATACAGAGAGTCCCT 785
QY 140 ValThrAspValSerValProLeuAspGIYArgGIuTrpSerLeuSerIlePheSerPh 160
|||||
Db 786 CGTCACTAGAGTGTCCGTCCTCCCGCGAGCGGCGCCAGTGGAGCTTCGATTTCCTCT 845
QY 160 ePromeThrArgThrAlaTyValAlaValAlaAsnValGIuAsnLysGIuMetSerLe 180
|||||
Db 846 TCCGATGTGAGAACCGCTTACGTCCGCGTAGCGAACGTGAGAACAGAGATGTGCT 905
QY 180 uAspValValAsnAspLeuIleGIuTrpLeuAsnAsnLeuAlaAspTrpArgTyValVa 200
|||||
Db 906 CGAGCTGTACAGCACTCATCGAGTGGCTCAACAATTCGCCACTGGCTTATGTCT 965
QY 200 lAspSerGIuGIuTrpIleAsnPhelThrAsnAspThrTrpTyrValArgIleArgVa 220
|||||
Db 966 TGACTGTGAACAGTGGATTAATTCACCAATGACACCAAGTACAGTACGCGACCTCCGCT 1025
QY 220 lLeuArgProThrTyrrAspValProAspProThrGIuGIuValArgThrValSerAs 240
|||||
Db 1026 TCTACGTCCAACTCAGACGTTCAGACCCGACAGAGGGCTGTGTGGCAGAGTCTAGA 1085
QY 240 pTyrArgLeuThrTyrrAsaIleThrCysGIuAlaAlaAsnMetProThrLeuValAspGI 260
|||||
Db 1086 CTACCGCTCACTTATAGAGGATGATGATGAGAGCCACATGCAACACTCGTCGACCA 1145
QY 260 nGlyPheTrpIleGlyGIuTyrrAlaLeuThrProThrSerLeuProGIuTyrrAspVa 280
|||||
Db 1146 AGGCTTTGGATCGCGGCGCAGTACGCTTCACCCCGCACTACCGCAGTACGACGT 1205
QY 280 lSerGIuAlaTyrrAlaLeuHisThrLeuThrPhalaArgProSerSerAlaAlaAla 300
|||||
Db 1206 CAGCGAGCGTACGCTGTGCACACTTGTGACCTTGCAGACCACTCCAGCGCTGCAC 1265
QY 300 uAlaPheValTrpAlaGlyLeuProGIuGIYThrAlaProAlaGIYThrProAlaTr 320
|||||
Db 1266 CGGCTTGTGTGGGCGAGTTGGCCACAGGGGAGGACGCGCTCGAGCACTCCAGGCTG 1325
QY 320 pGIuGIuAlaSerSerGIYGIYTyrrLeuThrTrpArgHisAsnGIYThrThrPheProAl 340
|||||
Db 1326 GGACAGAGGATCCCGGGGTGGCTACCTGACCTGGCGCACAAAGTACTCTTCCAGC 1385
QY 340 aGlySerValSerTyrrValLeuProGIuGIYPhalaLeuGIuArgTyrrAspProAsn 360
|||||
Db 1386 TGGCTCGTTAGTACGTTCTCCCTGAGGGTGTGGCTTGAGGCTACACCCGAACGA 1445
QY 360 pGIYSerTrpThrAspPheAlaSerAlaGIYAspThrValThrPheArgGIuValAlaVa 380
|||||
Db 1446 CGGCTTGTGAGCGCACTTCGTTCCGAGAGACACCGTCACTTCCGGGAGGTGGCGCT 1505
QY 380 lAspGIuValValThrAsnAsnProAlaGlyGIYGIYSerAlaProThrPheThrVa 400
|||||
Db 1506 CGAGAGAGTGTGTGACAAACACCCCGCGGGGCGGAGCGCCCGCCACTTCACCGT 1565
QY 400 lArgValProProSerAsnIleTyrrThrAsnThrValPhaArgAsnThrLeuGIuTh 420
|||||
Db 1566 GAGAGTCCCTTCAACAGCTTACACCAACACCGGTGTTAGAACGCTTTAGAGAC 1625
QY 420 rArgProSerSerArgArgLeuGIuLeuProMetProProAlaAspPheGIYGIuThrVa 440
|||||
Db 1626 TCGACCTCTCTGTAGGTGCGAAGTCCCATGCCACCTGTGACTTTTGACAGAGAGGT 1685

QY 440 lalaasnaasProLysIleGluInserLeuLysGluThrLeuGlyCysTyrLeuVa 460
 Db 1686 CGCACAACACCCGAGATGACGAGTCGCTTTAAAGAAACATTGGCTGCTATTGGT 1745
 QY 460 lHISerLysMeLArgAsnProValPheGlnLeuThrProLysSerPheGlyAlaVa 480
 Db 1746 CCACCTCCAAATGCGAAACCCGCTTTTCAGCTCAGCCGCGACGCTCTTGGCCGCT 1805
 QY 480 lSerPheAsnAsnProGlyTyrGluThrArgAspLeuProAspTyrThrGlyLea 500
 Db 1806 TTCCTTCACAAATCCGGGTTATGACGACACCGACCTCCGGACTACCTGGACTCCG 1865
 QY 500 gaSPserPheAspGlnAsnMeLserThrAlaAlaHisPheArgSerLeuSerHisSe 520
 Db 1866 TGACATATTGACCAACACATGTCACACCGCTGGCCCACTCCGCTCACTCCCACTC 1925
 QY 520 rGysSerIleValThrLysThrTyrGlnGlyTyrProLysGlyValThrAsnValAsnThr 540
 Db 1926 CTGCACTATCTGCTACTAGACCTACCAAGGCTTGGAAAGGCTCAGCAACGTCACAGCC 1985
 QY 540 oPheGlyGlnPheAlaHisAlaGlyLeuLeuLysAsnGluGluIleLeuCysLeuAlaAs 560
 Db 1986 TTTGGCCCAATTGCGCAGCGGGGCTCTCTCAAGAAATGAGAGATCTCTGCTCCGCGA 2045
 QY 560 pAspLeuAlaThrArgLeuThrGlyValTyrProLalThrAspAsnPheAlaAlaAlaVa 580
 Db 2046 CGACCTGGCCACCCGCTCTCAGAGTGTCTACCCCGCACATGACACATTCGCGCGCGCT 2105
 QY 580 lSerAlaPheAlaAlaAsnMeLuserSerValLeuLysSerGluAlaThrSerSerIl 600
 Db 2106 TTTCTGCTTTCGCGCGCAACCTGCTGCTCCGCTGGAATCGGAGGCAACGTCCTCCAT 2165
 QY 600 eileLysSerValGlyGluThrAlaValGlyAlaGlnSerGlyLeuAlaLysLeuPr 620
 Db 2166 CATCAAGTCCGTTGGGAGACTGCTCGCGCGGCTCACTCCGCGCTCCGGAAGCTACC 2225
 QY 620 oGlyLeuLeuMetSerValProGlyLysIleAlaAlaArgValArgAlaArgAlaAr 640
 Db 2226 CGGACTGCTAATGATGACACAGGAAGATTCGCGCGCTGCGCGCGCGCGCGAGCGCG 2285
 QY 640 gaArgAlaAlaAlaArgAlaAsn 647
 Db 2286 CCGCCCGCGCGCTGTCGCAAT 2307
 RESULT 8
 LOCUS SA43937 2448 bp RNA linear VRL 08-MAY-1993
 DEFINITION capsid protein [Nudarellia capensis omega virus, Genomic RNA, 2448 nt].
 ACCESSION S43937
 VERSION S43937.1 GI:255203
 KEYWORDS
 SOURCE Nudarellia capensis omega virus.
 ORGANISM Nudarellia capensis omega virus
 VIRUSES: ssRNA positive-strand viruses, no DNA stage; Tetraviridae; Omegatetravirus.
 REFERENCE 1 (bases 1 to 2448)
 AUTHORS Agrawal, D.K. and Johnson, J.E.
 TITLE Sequence and analysis of the capsid protein of Nudarellia capensis
 JOURNAL MEDLINE 92391097
 PUBMED 1519360
 REMARK
 Genbank staff at the National Library of Medicine created this entry [NCBI glibbsq 112777] from the original journal article.
 Map location: segment RNA2.
 FEATURES
 source 1..2448
 location/Qualifiers
 1..2448
 /organism="Nudarellia capensis omega virus"
 /db_xref="taxon:12541"
 367..2301
 gene

CDS
 /gene="capsid protein, CP"
 367..2301
 /gene="capsid protein, CP"
 /note="this sequence comes from Fig. 5; Protein sequence from Fig. 3 does not match conceptual translation; CP"
 /codon_start=1
 /product="capsid protein"
 /protein_id="AA023198.1"
 /db_xref="GI:255204"
 /translation="MDSNSASGKRSRRNRVRIANTVNVAPKORAGRARRANNID
 NVYAAOELGOSLDANVTPEFNVATMPERFARMGAKLIDDSIGWFKYLDPAAGAT
 ESARAVGERSKIPDLVYRSVYAEIREITNECPYSDATSLDQANSLSTSPMF
 RYAFYAVWVNDKEISLDVTNDLIVNNLWASRVVDSGQNFITSDDTFVRLRLV
 HPTYDLPDTBELRTRCSDYRLTYRSITCEANMPLVDGEWIGHYALPTATQNA
 VEGSGFHEFNTVRPIAGVTLTWASMPGASGDPAMIPDSIQFOMHGGFDA
 PGTIVTYTPRGVIMQYEDTTNENNGFANPDVLPFGQTGAAGTNAITITPAFTV
 LTIATTTTSAANVINFRNLDAETTTAASNSVEVLPPLFGOTAPNNPKTEOTLVKDTL
 GSIVYHSKRNRYFOLITPSSGALSTFRPGDRDLDPFGRGITDSIDVNNSTAVCH
 FSLSSCSITVTKTQGEVNTVNTPEQFAHSGLKNDTELCLADLADLATRLTGVC
 AIDNFAAVALAFNAANMLSVLSSEATTSVKELGNQATGLANQGLARLPGLIASIPGK
 IAAVRARDRRRRAARNNN"
 BASE COUNT 547 a 715 c 563 g 623 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 4,97e-135 Length: 2448
 Score: 2207.00 Matches: 428
 Percent Similarity: 76.07% Conservative: 71
 Best Local Similarity: 65.24% Mismatches: 139
 Query Match: 65.41% Indels: 18
 DB: 14 Gaps: 6
 US-09-677-653a-50 (1-647) x SA43937 (1-2448)
 QY 1 MetGlyAspAlaGlyValAlaSerGlnArgProHisAsnArgGlyThrArgAsnVal 20
 Db 352 CTGGGAACTTTTGAATGGACAGTAACAGCTCCGCGGCAAGGACGCGCAATGG 411
 QY 21 ArgValSerAlaAsnThrValThrValAsnGlyArgArgAsnGlnArg-----ArgArg 38
 Db 412 CGGATTGCAACCAACCGCTCAACGTCGCTCCGACCAACGCCAACACAGCGCGCCCA 471
 QY 39 ThrGlyArgGlnValSerProProAspAsnPheThrAlaAlaAlaGlnAspLeuAlaGln 58
 Db 472 GCCGTTCAAGTGCACCAACCAACATGTCACAGTGCAGCGACCTCAGCAATCGGCCAA 531
 QY 59 SerLeuAspAlaAsnThrValThrPheProAlaAsnIleSerSerMetProGluPheArg 78
 Db 532 TCTCTTGACGCAACAGCTTATACCTTCCCAACCAATGTTGCAACCATGCGCAATTCGT 591
 QY 79 AsnTrpAlaLysGlyLysIleAspLeuAspSerAspSerIleGlyTyrTyrPheLysTyr 98
 Db 592 AGCTGGCGCGGTGTAACATTCGACATCGACAGACATCCATTGGTTGATTTCAATAC 651
 QY 99 LeuAspProAlaGlyAlaThrGluSerAlaArgAlaValGlyLysTyrSerLysIlePro 118
 Db 652 CTGACCGCTGCTGGCGCCACGAGTCCGCTCGGCGGTGAGTACGCAAAATATACCT 711
 QY 119 AspGlyLeuValLysPheSerValAspAlaGluIleArgGluIleLysAsnGluGlyCys 138
 Db 712 GAGCGGCTGCTCAAAATTTCTGTGACGCGCGAATTAAGAGATATACATGAGGAGTGC 771
 QY 139 ProValAlaThrAspValSerValProLeuAspGlyArgGlnTrpSerLeuSerIlePhe 158
 Db 772 CTTACCGCTTCGACCTTCTATACCACTGACGCGCGCCCAATGAGCTTGTCTATCANC 831
 QY 159 SerPheProMetPheArgThrAlaTyrValAlaValAlaAsnValGluAsnLysGluMet 178
 Db 832 TCTTACCTATGTTTGGCACTGCTTACTTGTGTCGCGCAAGTGTATATAAGAGATTA 891
 QY 179 SerLeuAspValAlaAsnAspLeuIleGluTyrPheAsnAsnLeuAlaAspTrpArgTyr 198
 Db 892 TCCCTTGACGTACGAACAGATCTATCTGTTGCTTAAACAATCTCGTAGTTGGCGGTGAC 951

QY 199 ValValAspSerGluIntPrIleAsnPhrThAsnAspThrThrTyTyValAlaGlyIle 218
 |||||
 Db 952 GGTGGATGATGCGGACAGTGGTTCACCTTTTCGGACAGCCGACCTGGTTCGCCGAT 1011
 QY 219 ArgValLeuArgProThrTyArgAspValProAspProThrGluGluValAlaGlyThVal 238
 |||||
 Db 1012 CGGTGCTACACCCACATAGATCTTCGATCCGACCGAGGCGGTACTTCGTACGTCGC 1071
 QY 239 SerAspTyArgLeuThrTyTyLysAlaIleThrCysGluAlaAsnMetProThrLeuVal 258
 |||||
 Db 1072 TCGACACGACGACTTACTTATAGTCGATACCTGTGAGGCTAACATGCCACGCTGCT 1131
 QY 259 AspGlnGlyPheThrPileGlyGlyGlnTyArgAlaLeuThrProThrSerLeuProGlnTy 278
 |||||
 Db 1132 GACACAGGATTTGGATCGGAGACATTCGCTCTTACTCCGATCGATCCACACAGCAAT 1191
 QY 279 AspValSerGluAlaTyArgAlaLeuHisThrLeuThrPheAlaArgProSerSerAlaAla 298
 |||||
 Db 1192 GCGGTTGAGGCTTCGGGTTTGAGCACCCCTTTCACGCTCACCCGACCGGTTATCGACGC 1251
 QY 299 AlaLeuAlaPheValThrAlaGlyLeuProGlnGlyGlyThrAlaProAlaGlyThrPro 318
 |||||
 Db 1252 GGTGTACCTCAGCTCAGTGGCGTCATAGCCGCGGCTGGAGTCTCGCTCGGATGATCCT 1311
 QY 319 AlaTrpGluAlaSerSerGlyGlyTyArgLeuThrTrpArgHisAsnGlyThrPhe 338
 |||||
 Db 1312 GCTTGGATACCCGATTCACACGACAGCTC---CAGTGGGTATGAGGTTTTCAGCCT 1368
 QY 339 ProAlaGlySerValSerTyValLeuProGluGlyPheAlaLeuGluAlaGlyTyArgPro 358
 |||||
 Db 1369 CCCACCGGCTCATCTTACTATCTCCCGAGGCTACACCATGCAATCTTGGACACT 1428
 QY 359 AsnAspGlySerTyTrpThrAspPheAlaSerAlaGlyAspThrValThrPheArgGlnVal 378
 |||||
 Db 1429 ACCACGATGAATGAATGGTTGGCCATCCAGATGACGCTGTTACTTCGCGCA--- 1485
 QY 379 AlaValAspGluValValValThrAsnAsnProAlaGlyGly----- 393
 |||||
 Db 1486 -----ACTGGTGGCGCGCGGAACTAACCGCACCATTCAGATC 1524
 QY 394 SerAlaProThrPheThrValArgVal---ProProSerAsnAlaTyTrpAsnThrVal 412
 |||||
 Db 1525 ACAGCAGCTACTGTACTCTCACCATTCTGCCACGACCATCTCTGCTGAAGTAT 1564
 QY 413 ---PheArgAsnThrLeuLeuGluThrArgProSerSerArgArgLeuGluLeuProMet 431
 |||||
 Db 1585 AACTTCGGAATTTGGACGCTGAATAACCGCTCTTCTAATGCGAGTGAAGTCCCTCTG 1644
 QY 432 ProProAlaAspPheGlyGlnThrValAlaAsnAsnProIysIleGluGlnSerLeuLeu 451
 |||||
 Db 1645 CCACGCTTAACTTTGGTCAACCTGCGCGCGCAACCCGGAAGATGAACAGACTCTGCTC 1704
 QY 452 LysGluThrLeuGlyCysTyTyLeuValHisSerLysMetArgAsnProValPheGlnLeu 471
 |||||
 Db 1705 AAGGACACTTGGGTTCATACCTGTTCATACAGCAAGATGGCAACCGGTTTCCAACTC 1764
 QY 472 ThrProAlaSerSerPheGlyAlaValSerPheAsnAsnProGlyTyGluArgThrArg 491
 |||||
 Db 1765 ACAGCTGCACTTCTTTGGCGCATCAGTTTCACGAATCCCGGGTTTTCAGCGCAACCTC 1824
 QY 492 AspLeuProAspTyTrpArgTyLeuArgAspSerPheAspGlnAsnMetSerThrAlaVal 511
 |||||
 Db 1825 GATCTCTCTGTTTGGCGCATACGATGACTCCCTGATTCACCAATGCTCCACCGCTG 1884
 QY 512 AlaHisPheArgSerLeuSerHisSerCysSerIleValThrLysTyTrpGlnGlyTrp 531
 |||||
 Db 1885 TGCATTTTGTCTTGTCTAAGTCAATGTTCTATTCTTCATAGACGCTACCAAGTTGG 1944
 QY 532 GluGlyValThrAsnValAsnThrProPheGlyGlnPheAlaHisAlaGlyLeuLeuLys 551
 |||||
 Db 1945 GAAGGTGTACTAATGTTAACACCCCTTTGGTCAATTTGGTCTACACGCGGCTCTCAAG 2004

QY 552 AsnGluGluIleLeuCysLeuAlaAspAspLeuAlaThrArgLeuThrGlyValTyPro 571
 |||||
 Db 2005 AATGACACAGATTTCTTTCCTTTCGACAGACTGATACCCGCTTACAGCGCTTATGCT 2064
 QY 572 AlaThrAspAsnPheAlaAlaAlaValSerAlaPheAlaAlaAsnMetLeuSerSerVal 591
 |||||
 Db 2065 GCGACGATTAATTTTGTCTGCTGCGGCTCTGCTTTCGCCCAATATGCTCACTTCGCTC 2124
 QY 592 LeuLysSerGluAlaThrSerSerIleIleLysSerValGlyGluThrAlaValGlyAla 611
 |||||
 Db 2125 TTGAAGTCGGAAGCTACACTTCAGTATTAAGAGCTCGCATATCAAGCTACGACT 2184
 QY 612 AlaGlnSerGlyLeuAlaLysLeuProGlyLeuLeuMetSerValProGlyLysIleAla 631
 |||||
 Db 2185 GCCAACGAGGCGCTTGGCCCTTACCGGGCTTCTTCCCTCGATACCGGCGCAAAATTCCT 2244
 QY 632 AlaArgValArgAlaArgAlaArgAlaArgAlaAlaArgAlaAsn 647
 |||||
 Db 2245 GCACGTTCGCGCTACAGCTGACCGTCCGACGCTGCTCAGCAATGAAC 2292
 RESULT 9
 AF062037 2482 bp mRNA linear VRL 23-JUL-1999
 LOCUS
 DEFINITION
 complete cds.
 ACCESSION
 AF062037
 VERSION
 AF062037.1 GI:3861489
 KEYWORDS
 SOURCE
 Thossea asigna virus.
 ORGANISM
 Thossea asigna virus.
 Viruses; ssRNA positive-strand viruses, no DNA stage; Tetraviridae;
 unclassified Tetraviridae.
 REFERENCE
 1 (bases 1 to 2482)
 Pringle, F.M., Gordon, K.H., Hanzlik, T.N., Kalmakoff, J., Scott, P.D.
 and Ward, V.K.
 A novel capsid expression strategy for Thossea asigna virus
 (Tetraviridae)
 J. Gen. Virol. 80 (Pt 7), 1855-1863 (1999)
 JOURNAL
 MEDLINE
 99350019
 PUBMED
 10423156
 REFERENCE
 2 (bases 1 to 2482)
 Pringle, F.M., Kalmakoff, J. and Ward, V.K.
 AUTHORS
 Pringle, F.M., Kalmakoff, J. and Ward, V.K.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (29-APR-1998) Microbiology, University of Otago, 700
 Cumberland Street, PO Box 56, Dunedin 9001, New Zealand
 3 (bases 1 to 2482)
 Pringle, F.M., Kalmakoff, J. and Ward, V.K.
 REFERENCE
 AUTHORS
 Pringle, F.M., Kalmakoff, J. and Ward, V.K.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (12-NOV-1998) Microbiology, University of Otago, 700
 Cumberland Street, PO Box 56, Dunedin 9001, New Zealand
 REMARK
 Sequence update by submitter
 On Nov 13, 1998 this sequence version replaced gi:3859469.
 FEATURES
 location/Qualifiers
 1..2482
 /organism="Thossea asigna virus"
 /specific_host="Setothossea asigna"
 /db_xref="taxon:83810"
 /note="old name for host is Thossea asigna"
 1..2482
 /gene="TAV-CP"
 88..2361
 /gene="TAV-CP"
 /note="forms T-4 capsids with 240 copies of the large and
 small capsid proteins per capsid"
 /codon_start=1
 /evidence=not-experimental
 /product="capsid protein precursor"
 /protein_id="AAC97195.1"
 /db_xref="GI:3861490"
 /translation="MIVILKMLRSMNORYLEQNLKLIKIGKPLGFRSSGSGEOLVAA
 LANLIKPCVNSLRTHEHPSWSRGRTIOEVALMIKEIEELICVCPQLOEDTLISF
 ENPTGSSNLKQKSMFRRSRRLRRPRONGIVRREGSGSLITCGDVENGPPTVA
 GLVQEAIVGLINKATKTAIPKRRRRRRNRKGRIGNQPFSSISAGAPMKIRALRGLT

```

AMTERISPVSGMLETNLDPCGEKSTLDYGVDPGALPMVSGOFRETFTIRHPCAN
PMALPDGSMKTLGIIIRTNLMRTPELLADNMNAEISDESLDEVASENNYGLADA
SHNNVETSOAGLYWSIORWALIGVPSDIPYITDFRTSGFVSSNTPSLI
NCGMYVVOENFNTENKEYIPIHSESGEMPIGWSRLYSVAVALITYIOPGGAGONO
RTVTLNRSQSVSNSSVLPOVTOGYRRNNDYAVGLDLRQLTERDSGTYRRLORSA
DGTFTNIGPLSTTGGNGENFDSGLVLEGLDVLAGVNMNLTLPPTQSDLOQTP
KTCVFOIKEGGYVRODIIWQPYNMTPASRYAPVRVSNALITDLNNAVGTIRTDAD
SNYFALCHMTSLPLACAPFIKAEMRFEAVPGRNSWPGPFATSTPKDEVALTISRTV
MDLEPFAMPLAYNGWGLMFSKVISIYRIRPKALRTASNAVETVACVEDAQAVTKVY
ETVRGRGAMRMPTLQO"
88..552
mat_peptide
/gene="Tav-CP"
/product="unknown"
/note="not encapsidated in the mature virion"
553..2172
mat_peptide
/gene="Tav-CP"
/product="large capsid protein"
2173..2258
mat_peptide
/gene="Tav-CP"
/product="small capsid protein"
106
misc_feature
/gene="Tav-CP"
/note="alternate position for initiation methionine"
BASE COUNT      624 a      583 c      654 g      621 t
ORIGIN
Alignment Scores:
Pred. No.:      2,256-19      Length:      2482
Score:          436.00      Matches:      175
Percent Similarity: 42.37%      Conservative: 111
Best Local Similarity: 25.93%      Mismatches:  270
Query Match:    12.92%      Indels:      119
DB:             14          Gaps:        29
US-09-677-653a-50 (1-647) x AF062037 (1-2482)
QY      10  ATGPTCHIASn-----ArgarGlyThrArgAsn-Val---ArgValSerAlaAsnTh 26
DB      549  CGGCCCCACCGAGCGTGGCGCTGCCAGCGCGGTGGTGTCTCAACAAGGCAAC 608
QY      26  rValThrValAsnGlyArgArgAsnGlnArgArg-----ThGlyArgGlnVa 43
DB      609  CAAACCCGCTCTGTGTAAGAGCCGCGGCGCTAGACAAATGAGAAAGGAGATTGG 668
QY      43  lSerProProAspAsnPhetThAlaAlaAlaGlnAspLeuAlaGlnSerLeuAspAlaAs 63
DB      669  TATGCCATCAACCTTTTCATTTCCGCGC-----699
QY      63  nThrValThrPheProAlaAsnIlleSerSerMetProGluPheArgAsnTrpAlaLysG 83
DB      700  -----GGTGGCCCTATGCTCTAAATACCTGCCCTTCGGGCGCACCTTAGCGAT 746
QY      83  yLysIlleAspLeuAspSerAspSerIlleGlyTrpTrpPheLysTyLeuAspProAlaG 103
DB      747  GACCCGAACTATATCCCTCGTGGTGGGTGGTGAACCAACCTTGACCTTGCG 806
QY      103  yAlaThrGlnSerAlaArgAlaValAlGlyLysTyLeuSerLysIlleProAspGlyLeuVal 123
DB      807  GGAGTATAAATCCACTTTA-----GACTATGGAGAAAGTCCCTGATGGGGGCTATCC 857
QY      123  sPheSerValAspAlaGlnIlleArgGluLysArgGlnGluGlyCysProValValThrAs 143
DB      858  TATGCTCTGTGGCGCAATTTTCGAGAGACCTTCACTTATTCGGACCCCGGAGCGCAACC 917
QY      143  rValSerValProLeuAspGlyArgGlnTrpSerLeuSerIllePheSerPheProMetPh 163
DB      918  CATGTGGCCCTCTGTGAGGTTCATGTCAGCCCTCGGAGTAATAGAGACCAACCTGTG 977
QY      163  eArgThrAlaTyValAlaValAlaAsnValGluAsnLysGluMetSerLeuAspVala 183
DB      978  GGAACACTCGTTTATTATATATGCAATATGCAATATCTGAAATATCTGATGAATCCTT 1037
QY      183  lAsnAspLeuIlleGluTrpLeuAsnAsn-----LeuAlaAspTrpArgTyVa 199

```

```

DB      1038  AGACGAGGTGGCCCGGCTTTTCAACATATATGTGGAGACCTGGCCGAC----- 1086
QY      199  lValAspSerGluGlnTrpLysAsnPhetThAsnAspTrpThTyTyValArgLLea 219
DB      1087  -GGCTCCACCCSTAAATGCTGA--ACTAGTCAGGTCGGCTTATGTGCTCATC 1142
QY      219  gValLeuArgProThrTyLeuAspValProAspProThrGlnu---GlyLeuValArgThVa 238
DB      1143  AAGGTGACAGCTCTAATAGGTGTTCGCCCCCAGTATACAGGTTGTGTCGACAT 1202
QY      238  lSerAspTyArgLeuThrTyLysAlaIleThrCysGlnAlaAsnMetProThrLeuVa 258
DB      1203  AACGCACTCCGATAGTACGCGGTGGGTCTACTGTGCTCTCAACACACGCTGTGAT 1262
QY      258  lAspGlnGlyPheTrpIlleGlyLysLysLysLysLysLysLysLysLysLysLysLys 274
DB      1263  TAAATCAGGATAGTGTGTGTAGACACATTT-----AACCTAATATGAGAAATAGA 1316
QY      275  -LeuProGlnTyAspValSerGlnAlaTyAlaLeuHisThrLeuThrPheAlaAr 293
DB      1317  GGTATTTCTCACTGAAATCTGGGAGACCCCGTATAGCTGGCTCCGCTGATATAGTCT 1376
QY      293  gProSerSerAlaAlaAlaLeuAlaPheValTrpAlaGlyLeuProGlnGlyTyAl 313
DB      1377  TACAGCATCAACTGTCGCAATTAATATACAACTGGGGTT-----GGAGGT----- 1425
QY      313  aProAlaGlyThrProAlaTrpGluGlnAlaSerSerGlyGlyTyLeuThrTrpArgH 333
DB      1426  -----GCACAAATACAGAGACAGTACCTTAATAG 1457
QY      333  sAsnGlyThrThrPheProAlaGlySerValSerTyValLeuProGlnGlyPheAlaLe 353
DB      1458  GTACAGTCTCTGATATAAT-----ACTTACGCTTCCCGAC-----GTGAC 1499
QY      353  uGluArgTyAspProAsnAspGlySerTrpThrAspPheAlaSerAlaGlyAspThrVa 373
DB      1500  TCAAGGATCCGCTTAAATATGCG-----GACACATATGCTGAGGATAGCGCT 1550
QY      373  lThrPheArgGlnValAlaValAspGluValValAlaThrAsnAsnProAlaGlyLys 393
DB      1551  TCGCTTTCAG-----CTGACTGAGAGGATCTGTGCT-- 1584
QY      393  ySerAlaProThrPheThrValArgValProProSer-----AsnAlaTyThrAs 410
DB      1585  -----TACAGCTCCGCTTCAAGGTCGGCTGACGAGTACACCTTTACGA 1631
QY      410  n-----ThrValPheArgAsn-----415
DB      1632  TATAGTCCGCTTTCGACAAAGCAGACGCAATGCGGCTTTAAACAGATAGCGGCT 1691
QY      416  -ThrLeuLeuGluThrArgProSerSerArgTrgLeuGlu---LeuProMetProPr 433
DB      1692  TGGAGTCTTCTGGAGAGACAGCTGTAGCAGGCTGTAATATAGTCACTTACCTCC 1751
QY      433  oAlaAspPheGlyGlnThrValAlaAsnAsnProLysIlleGlnInsSerLeuLeuLysG 453
DB      1752  TTTTACTCATCAGATTTGATGACAGCAACGCAAGAGCTGTGCTTCCAAATAAAGA 1811
QY      453  uThrLeuGlyCysTyLeuValHisSerLysMetArgAsnProValPheGlnLeuThrPr 473
DB      1812  A-----GGTGGCTTTTATGATCGGCAAGGATTTTGGCAGCGGTGTACAAATATGACAC 1865
QY      473  oAlaSerSerPheGlyAlaValSerPheAsnAsnProGlyTyGlnAlaGlyThrArgAsp-- 492
DB      1866  TCGAGCCGCGATAGCTCTGTGCTTGTGTCAATAGCCATA-----ACTCTGATGTC 1919
QY      493  -LeuProAspTyTrpGlnTy---lLeArgAspSerPheAspGlnAsnMetSerThAlaVa 511
DB      1920  ATTAAATTCAGCAGAGGCGGATACATATAAGACACCCGCTGATGCTGATGCTTGGCCCT 1979
QY      511  lAlaHisPheArgSerLeuSerHisSerCysSerIlleValThrLysThrTyGlnGlyTr 531

```

Dh 1980 ATGTCAATGACCTCTTTGGCCCTTCGATGCGCTCGGTTATTAAGCGGAGATGGCCTT 2039

Qy 531 peltuglyValThrAsnValAsnThrProPheGlyGlnPheAlaHisAlaGlyLeuLeu 551

Dh 2040 TGAAGCGGCTCCGAGAGAAATTCCTCCGAGACCTTTGGACCTCCACCTCCTCCAA 2099

Qy 551 saangluGluIleLeuCysLeuAlaAspAspLeuAlaThrArgLeuThrGlyValTyrP 571

Dh 2100 GAGCAGAGCTTGGCTTACCTATTCACAGACTGTGATGAGACCTTGACCATTTCTATGCC 2159

Qy 571 oalathrAspAsnPheAlaAlaAlaValSerAlaPheAlaAlaAsnMetLeuSerSerVa 591

Dh 2160 CCGGCTTATATAGCTGGGGGTTGATGTTCTCGAAAGCTCTTATGTCGGAAGAT 2219

Qy 591 lleuYrsSerGluAlaThrSerSerIleIleYrsSerValGlyGluAlaValGlyAl 611

Dh 2220 ACCCAAGCCCTCAGAACCTCCAGC-----AATGTTCAGAGACTGTTCCAGGCTG 2270

Qy 611 a-----lAglInseryGlyLeuAlaValSerProGlyLeuLeuMetSerValProGl 628

Dh 2271 TGTAGAGAGATGCTCAGCGCCGCTCAG----- 2298

Qy 628 yLysIleAlaAlaArgValArgAlaArgAlaArgArg 642

Dh 2299 -AAAGTGTATGAAGACTGTCGTGGCGCGCAGAGACTATGAGA 2340

RESULT 10

AF461742 5698 bp ss-RNA linear VRL 20-FEB-2002

DEFINITION Euprosteria elaeasa virus, complete genome.

ACCESSION AF461742

VERSION AF461742.1 GI:18766345

KEYWORDS

SOURCE Euprosteria elaeasa virus.

ORGANISM Euprosteria elaeasa virus.

REFERENCE 1 (bases 1 to 5698)

AUTHORS Zeddam,J.-L.A., Pringle,F.M., Gordon,K.H., Ward,V.K., Luke,B.T., Goralenya,A.E. and Hanzlik,T.N.

TITLE Genome organization of Euprosteria elaeasa virus defines it as a member of a new group of insect RNA viruses

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 5698)

AUTHORS Zeddam,J.-L.A., Gordon,K.H. and Hanzlik,T.N.

TITLE Direct Submission

JOURNAL Submitted (20-DEC-2001) Insect Virology, IRD, Route de Montpellier, Saint-Christol-lès-Ales 30380, France

FEATURES

source

1..5698

/organism="Euprosteria elaeasa virus"

/db_xref="taxon:186165"

/note="Isolated from larvae of Euprosteria elaeasa, Dytar (Lepidoptera: Limacodidae)- a major defoliating pest of oil-palms in several Latin American countries"

95..3868

/codon_start=1

/product="putative replicase"

/protein_id="AAI78970.1"

/db_xref="GI:18766346"

/translation="MEASNPVIAPRSLSEAMLEERASVRODLSLRVLRGAPRTLAPDAPKGRDSDAQAASIVELRDVAEAGLEHKSIDPELOGEYVPCIIIGHDPIHPSGTNNKKKSVQROFTTSMSVDAIOLALREDPVLRDIALKTTTIDGSLAGPWRKRKRALPMDLCVPLRYOHIVDGLATIKKHPELFLAEKNKDRLEVEKLGKTRFYFSPHFLSALVSYSQSFKALTIMTEESTSYNAYGFSMTNGAEDLVIAWAGGEVKEPRLIACGDDTDIYRQKGRIRYICPDKDKDGSVDETTIAVVDYVVAHRSOYPLAKGFWLEGVKIMVEMATSPRLIDGTIYRKRKONGCLMGVGTLLFDIVKSAIAYKDMADOLLRBDLELKKKAKIKFEKKGHGYIKETGNPNAVTDPEIDELMTFOKFLGLQIKVYKADTDKCVFPLTPYSDWLSMWVTPARLGHKESYMRORTLEDRARGLVYGAAPKQARLEMGSVINTEPSEIVCMKVOEGGGRGAPSPYALTDNGNEFPVSDGFDPLATNTRLYSRDOTGDMRPVDFDVTLLAAYKKRLMDRIIRAVKEGVSLIMEVAVAETGIMNMPAEVAVLPKPSIOKPREFTVAENSTGSLVYKVKVDKSDSAGTLETFQKFMESVVKQIEELRAGQEKERTIIMGAALPAEIRKKMGVNDLRKTAEK

AAI78970.1:STNPVYLGHPDSESEGERAEVHKVYIKGHTPERKAREKPEVHD
TGEFVKITTPAGGCVPTQCEPOTPWYSICRSVSAGVOTVTRTERPGDHNRPVRA
VVLQRMRLKLPQWOBWLVTGSRILAKNELASDIALELVQERKQIPDATTSAED
ABLGDTIQLDRGRHVESLIGERFEVENVGISTKHQLDRGLMWOPASGYVVAAP
PKRTSGARSRTLASVKKSVVAVANDLRYRTAGEPFRDPVLSMLKKIKENKEARF
HSEEDLSREPEYORGPSSSFHEDSRADDESHADQHRVAVOOLSTKIDNKEARF
ABDARDREPSLDKREAGVPVPGSRGPRYSRLRPNPVVOOLEOVASLOAOITTTI
ACVRAPRPGCAKRSQNMNRGGESENAPORCGAGGSGGASQNKPRPKSM"
3336..5606

CDS

/codon_start=1

/product="capsid protein precursor"

/protein_id="AAI78971.1"

/db_xref="GI:18766347"

/translation="KIKIRRLSLMNQKILEANIRNIKEGHPLEAFMRTACREMSIMLA
NMIVCPTCNNSLRNKRERDWRGTLDLEENLAWIKRELRLCPDHDLLDQACGI
PPPESSNSKSVKHFHRSRLPESQLOQAGRSRLVYCGEENPGPPSVARGLQIE
VVAGLAKTGSPEAKRNGQARRRRRRYKCGTGNQNPAPVAGALATVLPALRGIE
MSEIISPSSVGMLETLDEDCGEYKSTLDYGVDPDGAIPMSVCGORFETTRHGPANP
MVAPLDGTMTWTIGIIRTNLMRPPILLIADMRNAEISDSLDVARSPNNVGEIELAS
HPNMMETLOEGLYMSIQRWTLALICVPPSPDVPSPYITDFRRTGSGFTYSSNTPSLN
QGMVVVAQFNPENTENKEIVPHSESGETPMGSRVSLSTQSIVTITIQGVGAQVTE
TVIANRPAVINSVLPAYVQYRELNGDTIYVGSRLYQTERDSGGTYVYLQNSTD
GTTFNIGPSPSTTIAFSGTSTASGVFLIEDVLAGRVNMLTLPEFTOSDLMQOTPK
TCVFOIKREGGFYVDVQVQVYNTMPASRYAVRFVSGAITDLSLNSVGTITRDADS
NGFALCHMTSILPIACAPRIKAEKMRPEAVGRNPSMGLAFTSTPPKDEVALTISRYM
DLEPFPMDLAYVNGWGLMPEFSKIINIYAKVXVLTASNVADVAVANCEVDQAATRYVE
SVGRRGANRDMTLOO"
3336..3785

/product="small putative peptide"

/note="not encapsitated in the mature virion"

mat_peptide 3786..5420

mat_peptide 3421..5603

BASE COUNT 1385 a 1357 c 1616 g 1340 t

ORIGIN

Alignment Scores:

Pred. No.: 5,02e-17 Length: 5698

Score: 407.50 Matches: 157

Percent Similarity: 39.69% Conservative: 97

Best Local Similarity: 24.53% Mismatches: 287

Query Match: 12,088 Indels: 99

DB: 14 Gaps: 22

US-09-677-653a-50 (1-647) x AF461742 (1-5698)

Qy 32 ArgArgAsnGlnArgArgThrGlyArgGlnValSerProProAspAsnPheThrAla 51

Dh 3876 AGGAGCGCGCGCTGCGCTTACACCGTTAAAGCAGCGGAATCCCAAT----- 3923

Qy 52 AlaAlaGlnAspLeuAlaGlnSerLeuAspAlaAsnThrValThrPheProAlaAsnIle 71

Dh 3924 -----CAGCCTTTTGG-----GTACCGCTGGCGCGCTCTA 3956

Qy 72 SerSerMetProGluPheArgAsnTrpAlaValGlyLysIleAspLeuAspSerAspSer 91

Dh 3957 ACGGGGGTCCCGCGCTTGGACACTCTCGCATGTGCGGACCACTTCACCTTGTGTCG 4016

Qy 92 ILGLITyrPyrThrLeuSerTyrIleAspProAlaGlyAlaThrGluSerAlaArgAlaVal 111

Dh 4017 GTAGGTTGGTGAAGACCAACTTGAGCCATGTGAGTAAAGTCGACCTTA----- 4070

Qy 112 GlyIleTyrSerIleProAspGlyLeuValIleValPheSerValAspAlaGluIleArg 131

Dh 4071 ---GACTATGCTAAGGTCCTATGATGCGCATCCCATGCGTGTGTCAGTTAAG 4127

Qy 132 GluIleTyrAsnGluIleGlyCysProValIleThrAspValSerValProLeuAspGlyArg 151

Dh 4128 GAGACTTTTCAGATCCGACATCCTGGGGCCCAACCAATGATGTCCTTTGATGACACC 4187

Qy 152 GlnTrpSerIleuSerIlePheSerPheProMetPheArgThrAlaIleValAlaValAla 171

Db 4188 ATGTGACACCTTGTATTAATCCGACTAATTTATGAGGACTCCGTTCTTGATTCGA 4247
 Oy 172 AANVALGLAASNLSEGLUMETSERLEUASPVAlVALASNPLeuLEGLUTRPLEUASN 191
 Db 4248 GCATATGAGGAATGCTGAATATCCGATTATCTCTAGATGAGGACTCGCTTTCAAC 4307
 Oy 192 AANLEUALAASPTPRATGTyTVALVALAASerSGLUGLINTPLeuASNPheThraSNA 211
 Db 4308 AATTTATGTCGGGAGATTGAATTAGGCTCAACCCGAATGGGTTGAGACCTTGACGAG 4367
 Oy 212 ThrThrTyTValValArgValLeuArgProThrTyTAspValProAspProThr 231
 Db 4368 GGCCTGATTTGG---TCCATCCAGCGGTGAGCGCTTAAATAGCGCTCCACCTCCAGT 4424
 Oy 232 GLU---GLYLeuValArgThrValSerAspTyArgLeuThrTyTysAlaIleThrcys 250
 Db 4425 GATACGGGGGTTTCCCTATATCACTGATTTCCGGATCACAGGTTCCGGATTTTACTG 4484
 Oy 251 GLUALAASmetProThrLeuValAspGlnGlyPheTrpIleGlyGlnTyTAlaLeu 270
 Db 4485 TCTTCCAAACACCCCATCTTATTAATCAAGGAGATGGTGGTGGCACAGTTT----- 4538
 Oy 271 ThrProThrSer-----LeuProGlnTyTAspValSerLys----- 282
 Db 4539 AACCCGAACACTGAGAACAGAGATAGTACCCATTTCTGAGCTGGGAGAGACTCCGATG 4598
 Oy 283 -----AlaTyTAlaLeuHisThr-----LeuThrPheAlaArgPro 294
 Db 4599 GCGTGGTCTCGGGTTATTTCCCTATCAACGTACAGATAGATTACGATTATTTACACA 4658
 Oy 295 SerSerAlaAlaAla-----LeuAlaPheValTyTAlaGlyLeuPro 308
 Db 4659 GGGGTGGAGGTGCCAGAAATCTTCAACAGTATGCCAAATAGAGGGCTCGTGAATT 4718
 Oy 309 GlnGlyGlyThAlaProAlaGlyThrProAlaTrpGlnAlaSerSergLysGlyTyT 328
 Db 4719 AATTAAGCTGGTTCTCCCTCTCACTCAAGTTTACCGT-----TTT 4760
 Oy 339 LeuThrTrpArgHisAsnGlyThrThrPheProAlaGly---SerValSerTyTValLeu 347
 Db 4761 TTA-----AATGGACACAGTATGACAGTTGGGAGACTCCCTCGCATAT----- 4802
 Oy 348 ProGlnGlyPheAlaLeuGlnArgTyTAspProAsnAspGlySerTrpThrAspPheAla 367
 Db 4803 -----CAGTTTACGGAGCGAGAC 4820
 Oy 368 SerAlaGlyAspThrValThrPheArgGlnValAlaValAspGlnValValThraSNA 387
 Db 4821 TCAGCGGGGTATACAGTCT---TATCTGCAAGGTCAACTGATGTGATACGACTTTACAAT 4877
 Oy 388 AsnProAlaGlyGlySerAlaProThrPheThrValArgValProProSerAsnAla 407
 Db 4878 -----ATTGGCCCGCGCTGACGACAAACGCGCCCATCTGGCACT 4919
 Oy 408 TyTThrAsnThr-----ValPheArgAsnThrLeuLeuGlnThraArgProSer 423
 Db 4920 TCACGCGCTTCTGGGGTGTTCATCTAGAACGTGATGTTTGGCCGACGCTGGAAT 4979
 Oy 424 SerArgAlaGlyLeuLeuProMetProProAlaAspPheGlyGlnThrValAlaAsnAsn 443
 Db 4980 ATG-----TTAACCTTACCGCCCTTTACCCAGCTGTGATCTGTCGAACAGACC 5027
 Oy 444 ProLysIleGlnGlnSerLeuLeuLysGlnThrLeuGlyCysTyTLeuValHisSerLys 463
 Db 5028 CCAAGAGAGTGTGTGTCCAGATCAAGAG-----GGTGGTTTATGTATGGCCAGAGAT 5081
 Oy 464 MetArgAsnProValPheGlnLeuThrProAlaSerSerPheGlyAlaValAserPheAsn 483
 Db 5082 GTTTGGCAACCGCTATATATGATGACCCAGCCGCTTATGCCCGCTGATTTGG 5141
 Oy 484 AsnProGlyTyTArgValArgThraArgAspLeuProAspTyTThcGly---IleArgAspSer 502
 Db 5142 TCG---GGAGCAATTAATTAGATAGCTGAATTTCTTGTATGAACATATCCGCGATACA 5198

Oy 503 PheAspGlnAsmetSerThraAlaValAlaHisPheArgSerLeuSerHisSerCysSer 522
 Db 5199 GCAGACTCTAATCATATGAGTTTGGCGCTGTCTACATGACCTCCATATACCTTGGCGTGGC 5258
 Oy 523 IleValThrLysThrTyTArgGlnGlyTyTArgGlnGlyValThraSNAValAsnThrProPheGly 542
 Db 5259 CCTTTATTAAGCTGAGATGCGCTTGAAGCGGTGCCAGTGGCAACTCTCCCTGGGG 5318
 Oy 543 GlnPheAlaHisAlaGlyLeuLeuLysAsnGlnGlnIleLeuCysLeuAlaAspSpleu 562
 Db 5319 CTTTGTGGCCACTCCACTCCCGCAAGATGAGTTGCTGATACCATATACAGCAGCGTC 5378
 Oy 563 AlaThrArgLeuThrGlyValTyTProAlaThrAspAsnPheAlaAlaValSerAla 582
 Db 5379 ATGGAAGCTTGAACCTTTGCCATGCCCTTCCCTATTAATAGGTGGGCTGATTTCTCA 5438
 Oy 583 PheAlaAlaAsmetLeuSerSerValLeuLysSerGlnAlaThrSerSerIleLeuLys 602
 Db 5439 AAAATCATCATATATAGTACGCAAGGTTCCGAGAGTCTCCGACAGCAGCAACGTTGCT 5498
 Oy 603 SerValGlyGlnThAlaValAlaGlnSerGlyLeuAlaLysLeuProGlyLeu 622
 Db 5499 GATGTTTGTCTAACTGCTGGAGAGAGCCCAAGCTGGTGCACAC----- 5543
 Oy 623 LeuMetSerValProGlyLysIleAlaAlaArgValArgAlaArgAlaArgArg 642
 Db 5544 -----CGAGTTATAGAGATGTGGGGGGCGCGAGGGCTATGCCA 5585

RESULT 11
 LOCUS AF102884 6625 bp RNA linear VRL 30-MAY-2001
 DEFINITION Nudaurelia capensis beta virus complete genome.
 ACCESSION AF102884
 VERSION AF102884.1 GI:4028586
 KEYWORDS
 SOURCE Nudaurelia capensis beta virus.
 ORGANISM Nudaurelia capensis beta virus
 Viruses: ssRNA positive-strand viruses, no DNA stage; Tetraviridae; Betatraviruses.
 REFERENCE 1 (bases 1 to 6625)
 AUTHORS Gordon,K.H., Williams,M.R., Hendry,D.A. and Hanzlik,T.N.
 TITLE Sequence of the genomic RNA of nudaurelia beta virus (Tetraviridae) defines a novel virus genome organization
 JOURNAL Virology 258 (1), 42-53 (1999)
 MEDLINE 99263183
 PUBMED 10329566
 REFERENCE 2 (bases 1 to 6625)
 AUTHORS Gordon,K.H.J., Williams,M.R., Hendry,D.A. and Hanzlik,T.N.
 TITLE Direct Submission
 JOURNAL Submitted (30-OCT-1998) CSIRO Entomology, GPO Box 1700, Canberra, ACT 2611, Australia
 FEATURES
 source location/Qualifiers
 1. 6625
 /organism="Nudaurelia capensis beta virus"
 /db_xref="taxon:85652"
 93..5870
 /note="replicase"
 /codon_start=1
 /product="RNA-dependent RNA polymerase"
 /protein_id="AAC97509.1"
 /db_xref="gi:4028587"
 /translation="MEDASKOLRVLDNOERKAFAOLFIVASVETLEDAOEKYEGMF
 RSGTRPSTHKLALIDLRVAEKDRLRHVKVNPVLEIGSVESYRAVAVOTRDERVIG
 CTFSRDARDNLRNKNTIGYEATRTIGDPDALIAGTIPDTFCYVGFNSCEQSLATA
 CHSLYPPDESWSIMDVAKGMLHGTHYIVAMHLPVELLTITDNDNIEGISTIFEET
 GALPCTKRRAKIFSGYNDPGSAVYVADHNAWGLKHKGVDPYGFSLIDIQGFFGMH
 TKLITRGHSSGSTITVPLSLGLIWNPNVVKIMYPRKAKPEYIYDKKYEGVCV
 YVGRVOSGSKSTILAEIYQITRILRLITNGVHEKWTWIIADODIERLAKYSIMFK
 NVERAVSEKALPRAOKCKSEKQALLPYMVRIRIANPFOKFOIDEVYRRKYIECK
 AQPWTHADKYVNCETKRINPVAIVGKRNHLATGTGRELREIPLSANEPDRKAKM
 HSAHADLDIYAEGLIDSAKAAAGKOSLITLQDAFOVLKRTCECCNTEIYEWG
 PGSGKSRNAAPRPADLOGVLVLCAPRTLRILDALDESIVHPSRVCYTHNMLHVAKES

CDS

GNRPDVIVIDEAETTPACVGTMMHSPSRIVCLGDPHQIGYIDFSRDKDLKPFS
 IIAECRRFRNTYKCPQDVLNLPFLKTLVDAISFKSLTSIRLTPARSVTRRRA
 AQTIDODPHESEPTTAHEPQARRDVVHAGTLEPEALILEKHIVNVALTRHNA
 LYIRESEKGEVLPSLMPSPSTYRCPVDKQWDPVAVERENSSSPCSHHIGA
 LITIDELGLDTKGVRFESAVFLAHRRVLDLGGSDPDPTMYOTNLNGTIT
 NTKDQOALHTLVGRYARKINSSREDAEDVKRITARLKEIMPETAPEQDSFA
 DAMOKIAERHGVVDIIDFWSNEGQRIYHKLQGVMPPTKLGKLOQSAHEKANI
 ALSAMVRIIDODMSTSEKIFPANGGSDRDTMSIIEERLOEKREPSIDKEEDYSR
 WYSLIVSWRCDRGCGDEHLIEYEERKSRSTSSIRLSRGSDVSMGSLDGVMTART
 PASGLMALFVGVDTLAKGDVFLAGNLLYIDAEELRSGSTLANNKIETAYVSE
 IGFIVSOAAVTVADVRLATRTYGRSTKNADILAKYIATADCKLERSPTRLMTAIN
 CATLGTGSKECINLMDALDAFGHTKMSDLHDPGVMTAPKVDERYSGDDGCR
 ADKTRKEQEPGCPQPOQASTQESGTSRSTSDQPRDGRTPRREPPGQPR
 SDTRREGVAKSDGESHSDIRGMDSLRSRGRIODEPGRRENSRRDTSVNRSTSR
 DRGROIPTGEFEYDATABMDLASTSDASVVLASVYVHHHOOHSGSPDSGCVRE
 RLEODGDLDSDVPKLGASREYVLRHPRDSADGRTPTSTGCIYRTRLPDIERH
 SVLQRTHSRESGSDRAVPTGQRTPEGGHSSRHPNGRNVATARFAELHIDDD
 RGPGRGRSNPATHGVDAGVAGVAPDEPDILRRKHNNHHHAATRYDDNVA
 IHSQORDGRHDSGATVAVRSEFGLTESAGHQLNDSTHEHPNDAGNAODHSV
 PIORNGGLYAPEGVPTRVNRNGDYLMTGAMEDTDNCGRLPPGMMWPGYHROOLR
 DRCRDDRVVYINRLLGQVPTLRSDTGGGEPGLGRDCTSGRGASGSLDPSA
 PIRIETTORIRGPIRDGQDHSPTSLCAISRSSECGDLHRENDRCSELEHJLGE
 AAKSETCWRNRSRSPQSCPHREP

CDS

4039..5877
 /note="coat protein; forms T=4 capsids with 240 copies of
 protein per particle"

/codon_start=1

/product="capsid protein"

/protein_id="AAC97510.1"

/db_xref="GI:4028588"

/translation="MDANYOIRPARNNPSQGNORNNNNRRRRRGLKLPVAVIT
 SPGMAEPANHANTRYNRGTRRGIRLOAMSPMASTSEAWTLDLDDGKTSID
 DKIRPDGAIPQSTCCQFRGTGVARRYRGLSTLPLDGGTWPILVHLPFRFRHLEIT
 TTSNTEVEVNTADLDLFAFANDMNRDWTATYPSMVQVNVNFTVMTALDTPPT
 OLQSGLEVTNADLSSGTATFNAPLNVQVAVIAPDPDEHOKENDIYAGTQVOT
 GGTLOQSGGPNYTLTMTIGDQVEFGAAILPLPTVSMGPEEGQVLPOTANLTFPG
 NTITTTTTPGSPVTGMQFASNGDVTVDAGALVAFGALDSELDLNDISIK
 IPTNNOMQAMPKTIOFOLNETGFWALDRAFPQVFPFMTATGAPRMRKPRPTV
 VDHRAIGLODTIDSNPAIGVAMTGMTSTVPYPRKVTNRREALTAIESPMPGPRSA
 TTPKDDVALTVARMTDHLDFATPERYNGFALFAVAVATLQIDIPRYVSAGVANAV
 TDCIESATESVSNSTSERORRRARVGIIARGLNLRIGNLST"

5644..5649

/note="cleavage site for coat protein after assembly"

6514..6625

/note="tRNA-like structure; valine anticodon; no
 pseudo-knot in the aminoacyl stem"

BASE COUNT 1696 a 1800 c 1783 g 1346 t

ORIGIN

misc_feature

5644..5649

5644..5649

5644..5649

5644..5649

5644..5649

5644..5649

5644..5649

5644..5649

5644..5649

5644..5649

5644..5649

5644..5649

5644..5649

5644..5649

5644..5649

5644..5649

5644..5649

5644..5649

5644..5649

5644..5649

5644..5649

5644..5649

5644..5649

5644..5649

5644..5649

5644..5649

5644..5649

5644..5649

5644..5649

5644..5649

5644..5649

5644..5649

5644..5649

5644..5649

5644..5649

5644..5649

5644..5649

4213 ---CGGTCACACCGCGTCGACACCGCGTGAAGGCGGTTAAGCAAGCATGATGAGAGT 4269
 86 AspleuAsperispsertileglttpttpttphelysytyleuAspProalaglyalaThr 105
 4270 CCATGCGACGACATGATGAGGATGATTCACGACTATGTAGACCGGAGGAGATAC 4329
 106 GluserlAarglAvala1GlytYserlYsileProasplyleuVallyspheSer 125
 4330 AAGACGAGCTGACGACGAGG-----AAATTCACGACGCGCGGATACCTGACTCA 4380
 126 ValAspAlaIuileAargIuileTyraSngluGluCysProValaValThrAspValSer 145
 4381 ACATGCGGTCATATTCAGAGGACCGTGGCGCCAGATACCGGAGATGATTCAGACG 4440
 146 ValProleuAspGlyAargIntpttpttserleuSerilepheserPrometpheArgThr 165
 4441 CTACCGGTGATGCGGCGGACCTGCTCTACTGTATGATCTCCGTTCTTCAGGAT 4500
 166 AlatyValaIaValaIaAsnValaIaAsnValaIaAsnValaIaAsnValaIaAsn 184
 4501 CCGTGTGTCATACACACGACGACGACGACGACGACGACGACGACGACGACGACGAC 4554
 185 Aspleu-----ilegttpttpttpttpttpttpttpttpttpttpttpttpttptt 200
 4555 GATCTGATGCGTTCGCGACGATTCG-----AACACAGAGAGAGAGAGAGAGAGAGAG 4611
 201 AspserleuIntpttpttpttpttpttpttpttpttpttpttpttpttpttpttpttptt 220
 4612 ---TACCAAGTTGGGCGCAAGTCGGGAC-----GTGTTTACAGATGTCGTC----- 4656
 221 LeuArgProThr-----TyrAspValProAspProThrGlu---GlyLeuValArg 236
 4657 -----CCGACCGAGAGCGCTGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4710
 237 ThrValSerAspTyrAargleuthrTyrrylsAlaIlethrcysgluAlaAsnMetProThr 256
 4711 TTACTGAGAGTTACCGTTCGACATCAGAGCGGCTCAGAGGATCTCAACGACCCACT 4770
 257 LeuValAspIntnglythetprrilegltylegltyleuValaIeuthrProthrSerleuPro 276
 4771 CTCTGATACAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4824
 277 GlutyrAspAlaSerGluAlaIleuAlaIleuAlaIleuAlaIleuAlaIleuAlaIleu 290
 4825 CAGAGGAG 4884
 291 ---PheAlaArgProSerSerAlaIleAlaIleAlaIleAlaIleAlaIleAlaIleAlaIle 309
 4885 GCGGTCACAGGCGGACCTACATGACGATGACG-----ATCGGAGAGAGAGAGAGAGAG 4941
 310 GlyGlyThrAla---ProAlaGlyThrProAlaIleProAlaIleAlaIleAlaIleAlaIle 328
 4942 GGGGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 5001
 329 LeuThrArgHisAsnnglythThrPheProAlaGly-----SerValSerTyr 345
 5002 CTGGTGTTCAGACTGCGAACCTGACATTGACCTCGGAAACACATACCATACGACAC 5061
 346 ValLeuProGluGlyPheAlaIleuGluArgTyraSproAsnAspGlySerTptThrAsp 365
 5062 ACCTGTCACACAGGCGGCGG-----ACGGAGATGTGCAATTC 5100
 366 PheAlaSerAlaGly---AspThrValThrPheArgIleValaIaValaIaAspGluValaI 384
 5101 ACACCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 5136
 385 ValThrAsnAspProAlaGlyGlyGlySerAlaProThrPheThrValaIaValaIaProPro 404
 5137 -----GCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 5178
 405 SerAsnAlaIleTyrrAsnThrValaIlePheArgAsnThrLeuGluIuThrArgProSerSer 424
 5179 TCGGAAGTGAATGTGAGGAGATC----- 5202

Alignment Scores:

pred. No.: 1,45e-15 Length: 6625
 Score: 386.50 Matches: 174
 Percent Similarity: 38.81% Conservative: 86
 Best Local Similarity: 25.97% Mismatches: 295
 Query Match: 11.46% Indels: 115
 DB: 14 Gaps: 31

US-09-677-653a-50 (1-647) x AF102884 (1-6625)

15 AArgGlyThraGAsnValaIaSerAlaAsnThraValaIaSngIaArgArgAsn 34
 4060 AGACCCGCGAGAAACACCGGACGACGAGCGGCGCAACACACACACACACGCT 4119
 35 GlnArgArgArgThraGlnValaIaSerPro-----ProAspAsnPheThrAla 51
 4120 CGACGCGAGGAGCGGCGGCTTAAACACTCCCGCTAGTCCGACCG-----ATTACGACG 4173
 52 AlaAlaGlnAspLeuAlaGlnSerIleuAspAlaAsnThraValaIaPheProAlaAsnIle 71
 4174 CCGGCGCGAGTGGCCGACACCGGACACGCGGACACAC----- 4212
 72 SerSerMetProGluPheArgAsnThrPalaIleGlyGly-----LysIle 85


```

QY 425 ArgArgLeuGluLeuProMetProProAlaAspPheGlyGlnThrValAlaAsnAspPro 444
DB 5203 -----ACTCATTCAGATTCACCAACAGACATGACCAATGATGCGGACAGCC 5256
QY 445 LysIleGluGlnSerLeuLeuGlnThrLeuGlyCysArgLeuValHisSerLysMet 464
DB 5257 AAGACCATTCAGTCCACATCAACAGCAAGCAAGGCTTTATATG---CCCTGAGGGCG 5313
QY 465 ArgAsnProValPheGlnLeuThrProAlaSerSerPheGlyValAlaSerPheAsnAsn 484
DB 5314 TTCACACCCGTTGCAAAATGACAAATGCGACGCTTATGACCGGTCGCGATGAAGACA 5373
QY 485 ProGlyTyrGlnArgThrArgAspLeuProAspTyr-----ThrGlyLeuArg 500
DB 5374 CCG-----AGGACA---ACTGTGTAGATTACCAACCGGCGCATTTGGTGCATCCAG 5421
QY 501 AspSerPheAspGlnAsnMetSerThrAlaValAlaHisPheArgSerLeuSerHisSer 520
DB 5422 GATACCATCGACAGCAACTCTCGCATAGGCGTTGCGCGCATGACCGGTATGTCTACATCA 5481
QY 521 CysSerIleValThrArgThrArgGlnGlyTyrGlnGlyValThrAsnValAsnThrPro 540
DB 5482 ACCGTACCTTACTTCAAGGTGTTCGACCGCTTCGAAGCGATTACCGCGAGGGAGCCCT 5541
QY 541 PheGlyGlnPheAlaHisAlaGlyLeuLeuLysAsnGluGlnIleLeuGlyLeuAlaAsp 560
DB 5542 TGGGGCCCTTGGCCAGTGCAGACACTCGAAGAGACAGCTGGCGGTATACAGTGCCTCA 5601
QY 561 AspLeuAlaThrArgLeuThrGlyValTyrProAlaThrAspAsnPheAlaAlaVal 580
DB 5602 ACTTGGACCGATTCGACCCATTCGCATCCCGAAGCATACACGATTCGGGGCCCTA 5661
QY 581 SerAlaPheAlaAlaAsnMetLeuSerSerValLeuLysSerGlnAlaThrSerSerIle 600
DB 5662 TTCGCGATGTGGCCAGACCATAGCCCAAGATACCTGCG-----TAT 5703
QY 601 IleYsSerValIleGlyThrAlaValAlaGlyAlaGlnSerGlyLeuAlaLysLeuPro 620
DB 5704 GTGGCATCA-----GCAGCGGAGTGGCGAATGCGGTGACGCGACTGCATAGAG 5751
QY 621 GlyLeuLeuMetSerValProGlyLysIleAlaAlaArgValArgAlaArgAlaArg 640
DB 5752 ACCGCGACCGAGAGTGTCTCGAATTCACCTCACCTCGAGAGCGCGCAAGACGAGAGA 5811
QY 641 Arg-----ArgAlaAlaArg 645
DB 5812 CGTGTGGCGGAATCGCTCGAGAGCCCGC 5841

RESULT 12
ARI59969 6534 bp DNA linear PART 17-OCT-2001
LOCUS ARI59969 Sequence 4 from patent US 6251654.
DEFINITION ARI59969
ACCESSION ARI59969
VERSION ARI59969.1 GI:16222854
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 6534)
AUTHORS Gordon,K.Heinrich. and Hanzlik,T.Nelson.
TITLE Modified small RNA viruses
JOURNAL Patent: US 6251654-A 4 26-JUN-2001;
FEATURES
Location/Qualifiers
source 1..6534
/organism="unknown"
BASE COUNT 1667 a 1782 c 1747 g 1338 t
ORIGIN

```

Alignment Scores: 1.56e-09 Length: 6534
 Pred. No.: 294.00 Matches: 184
 Score: 37.50% Conservative: 74
 Percent Similarity:

```

Best Local Similarity: 26.74% Mismatches: 281
Query Match: 8.71% Indels: 151
DB: 6 Gaps: 35

US-09-677-653A-50 (1-647) x ARI59969 (1-6534)
QY 2 GlyAspAlaGlyValAlaSerGlnArgProHisAsnArgGlyThrArgAsnValArg 21
DB 4012 GGGCAACGACGGCGCGGCAACACAAACAAACAGCTCGACGAGAGCGCGGTCTAAAA 4071
QY 22 ValSerAlaAsnThrValThrValAsnGlyArgArgAsnGlnArgArgThrGlyArg 41
DB 4072 CTTCCCGCCCTAGTCCACCGATT----- 4095
QY 42 GlnValSerProProAspAsnPheThrAlaAlaAlaGlnAspLeuAlaGlnSerLeuAsp 61
DB 4096 ---ACGACCGCGCATGATGCGACACCGCGAAC-----CAC 4128
QY 62 AlaAsnThrValThrPheProAlaAsnIleSerSerMetProGluPheArgAsnThrAla 81
DB 4129 GCGAACACCGGGGTCAA-CCGGGTG-----GACACGCGTGAAGGGGCT 4172
QY 82 LysGlyLysIleAspLeuAsp-----SerAspSerIleGlyTyrPyrPheLysTyr 98
DB 4173 AAGGCAAGCGATGATGAGAGATCCCATGGCGACGACATCAAGGCGATTCACGACTAT 4232
QY 99 LeuAspProAlaGlyAlaThrGlnSerAlaArgAlaValGlyGlnTyrSerLysIlePro 118
DB 4233 CTAGACCCGAGCGAGAAATACAAAGACGAGCTGACGACGGG-----AAATTTCC 4283
QY 119 AspGlyLeuValLysPheSerValAspAlaGlnIleArgGlnIleTyrAsnGlnGlyCys 138
DB 4284 GACGGCGCGATACCTCAGTCAACATCGCGGTCAATTCGAGGAGCGCGTGGCGCCAGATAC 4343
QY 139 ProValValThrAspValSerValProLeuAspGlyArgGlnThrSerLeuSerIlePhe 158
DB 4344 CCGGACGTGAATTCAGACGCTACCGCTGAGTGGGAGGCGGACCTGCTTACTAGAGATG 4403
QY 159 SerPheProMetPheArgThrAlaTyrAlaValAlaAlaAsnValGlnLysnGlnMet 178
DB 4404 CATCTCCGTTCTTCAGCGCATCGGTGTTGTTTCATCACCAACCGACGACGAGAA---- 4460
QY 179 SerLeuAspValValAsn---AspLeu-----IleGluThrLeuAsnAsnLeu 193
DB 4461 ---GTGAGATGACGAAACCGCGATCTGATGCGTGGCGAAGCATTTG---AACACAGG 4514
QY 194 AlaAspTyrArgTyrValValAspSerGlnGlnThrIleAsnPheThrAsnAspThrThr 213
DB 4515 ACGGACTGGACCGACGACGAG---TACCCAGTTGGCGCAAGTCCGGAAC-----GTG 4565
QY 214 TyrTyrValArgIleArgValLeuArgProThr-----TyrAspValProAspPro 230
DB 4566 TTTTACATGTGTC-----CCGACCGAAGCGCTGACGAGCATACACCCCGC 4613
QY 231 ThrGlu---GlyLeuValArgThrValSerAspTyrArgLeuThrTyrLysAlaIleThr 249
DB 4614 ACTCAACTGGGTATCAAGGTTACTCGAGAGTACTCGTGCATCGACGAGCGGCGTCAAA 4673
QY 250 CysGluAlaAsnMetProThrLeuValAspGlnGlyPheTrpIleGlyGlnTyrAla 269
DB 4674 GCGTACTTACGACGACCCACTCGTAATCGAGGAGTGGCGGTGTCGCGCATTC--- 4730
QY 270 LeuThrProThrSerLeuProGlnTyrAspValSerGlnAlaThrAlaLeuHisThrLeu 289
DB 4731 ---CAACCGACAAAGACACACAGAGAGAACCGGACATGTACTCCGGACACCAA 4787
QY 290 Thr-----PheAlaArgProSerSerAlaAlaAlaLeuAlaPhe 302
DB 4788 ACGGGGGAAGCTTACAGCTCGCGGTTCAGGGCGCAATCAACATTCGATGAGC--- 4844
QY 303 ValThrAlaGlyLeuProGlnGlyThrAla---ProAlaGlyThrProAlaThrGln 321
DB 4845 ATCGGGGACCAAGTGCAGTTGCGGGGCGACGAAATCCGCTACCCACGCGTGCATGGGG 4904

```

```

QY 322 GlnAlaSerSerGlyTyrLeuThrProArgHisAsnGlyThrThrProAlaGly 341
DB 4905 CCGATCCCGGAGTGGGGAGCTGGTGTCCAGACTGCGAATCGATTCAGCTCGGA 4964
QY 342 -----SerValSerTyrValLeuProGluGlyPheAlaLeuGluArgTyrAspPro 358
DB 4965 AACACAAATCCATCCAGACAGCTGCCACCAAGGCGGTG----- 5006
QY 359 AsnAspGlySerTyrThrAspPheAlaSerAlaGly---AspThrValThrPheArgGln 377
DB 5007 ---ACGGGAATGGCAATTCACAGCCAGCAGGAGCAGACCGTGACC----- 5054
QY 378 ValAlaValAspLeuValValThrAsnAsnProAlaGlyGlySerAlaProThr 397
DB 5055 -----GTGGAC-----GCGGAGACG----- 5069
QY 398 PheThrValArgValProProSerAsnAlaTyrThrAsnThrValPheArgAsnThrLeu 417
DB 5070 ---ACTGTACGCGTTCGGACCGAATTTGGACGCGCTCGGAATCGAATTCGACGACATCAA 5126
QY 418 LeuGluThrArgProSerSerArgArgLeuGluLeuProMetProProAlaAspPheGly 437
DB 5127 CTCAAATCAAGATTCACCAAC-----GACATGAACCAAAATGATGACAGGC 5171
QY 438 GlnThrValAlaAsnAsnProGlyIleGluGlnSerLeuLeuGly---GluThrLeuGlyCys 457
DB 5172 -----AACGCCAAGACCAATTCAGTCCCAATCAAGCAACGAGGCGATATATATATG 5219
QY 457 sTyrLeuValHisSerLysMetArgAsnProValPheGlnLeuThrProAlaSerSerPhe 477
DB 5220 CCG-----AGCGGTCAATCCGTTGTCGAATATGCAAT---GCGAGCTCTTA 5263
QY 477 eGlyAlaValSerPheAsnAsnProGlyTyrGluArgThrArgAsp---LeuProAspT 496
DB 5264 TGGACGCGTCGAT-----GAAAGACACGAGACCAACTGTGTAGATTT 5305
QY 496 yTrhArgLys-----IleArgAspSerPheAspGlnAsnMetSer----- 508
DB 5306 AACCCGGCAATTTGGTCCACTCCAGATACATCGACACCACTTCGATAGCTGGCC 5365
QY 509 -----ThrAlaValAlaHis-PheArgSerLeuSerHisSerCysSerIleValThr 525
DB 5366 GTGATGACCGGTATGTCATCATCAACCTACCTACTTTCGACAGCTTCGACGCTTCG 5425
QY 526 LysThrTyrGln-GlyTyrPgluGlyValThrAsnValAsnThrProPheGluGlnPheAl 545
DB 5426 AACGCAATCCGGGAGGAGGAG-----CCTTGGGGCCCTTCGCG 5464
QY 545 aHisAlaGlyLeuLeuLysAsnGluGluIleLeuGlyLeuAlaAspSerLeuAlaThrArg 565
DB 5465 TAGTGGACACCTCCGAGGAGCGAGTGGCGCTAACAGTGGCTGCAACTGGACCGATCT 5524
QY 565 GLeuThrGlyValTyrProAlaThrAspAsnPheAlaAlaValSerAlaPheAlaAl 585
DB 5525 GCAACCATTCGCAATACCCGGAAGATCAACGATTCGGGGCCCTATTCGAGATGGCG 5584
QY 585 aAsnMetLeuSerSerValLeuLysSerGluAlaThrSerSerIleIleLysSerValG 605
DB 5585 CAAGACCATAGCCAGATACCTCGC-----TATGTGCGATCA----- 5621
QY 605 yGluThrAlaValAlaGlnInsSerGlyLeuAlaLysLeuProGlyLeuLeuMetSe 625
DB 5622 -----GCGACCGGAGTGGCGAATGGTGACGCGACTCGCATAGACCGCGAGAG 5674
QY 625 rValProGlyLysIleAlaAlaArgValAlaArgAlaArgAlaArgArg----- 641
DB 5675 TGTAGCTCGAATTCACCTCGGAGAGCGGCAACGAAGCGGAGCTGTGTGCGGAT 5724
QY 642 -----ArgAlaAlaArg 645
DB 5735 CGCTCGAGGAGCCCGC 5750

```

```

RESULT 13
LOCUS AX028953 6536 bp DNA linear PAT 16-SEP-2000
DEFINITION Sequence 4 from Patent WO9746666.
ACCESSION AX028953
VERSION AX028953.1 GI:10189994
KEYWORDS
SOURCE
ORGANISM Nudarellia capensis beta virus.
Nudarellia capensis beta virus
viruses; ssRNA positive-strand viruses, no DNA stage; Tetraviridae;
Betatetraviruses.
REFERENCE
AUTHORS Gordon, K.H. and Hanzlik, T.N.
TITLE Modified small rna viruses
JOURNAL Patent: WO 974666-A 4 11-DEC-1997;
COMM SCIENT IND RES ORG (AU) ; GORDON KARL HEINRICH (AU) ; HANZLIK
TERRY NELSON (AU)
FEATURES
source 1..6536
location/Qualifiers
BASE COUNT 1667 a 1782 c 1747 g 1338 t 2 others
ORIGIN
/d_basiseq="Nudarellia capensis beta virus"
/d_basiseq="taxon:85652"
Alignment Scores:
Pred. No.: 1.56e-09 Length: 6536
Score: 294.00 Matches: 184
Percent Similarity: 37.50% Conservative: 74
Best Local Similarity: 26.74% Mismatches: 281
Query Match: 8.71% Indels: 151
Gaps: 35
US-09-677-653a-50 (1-647) x AX028953 (1-6536)
QY 2 GlnAspAlaGlyValAlaSerGlnArgProHisAsnArgGlyThrArgAsnValArg 21
DB 4014 GGGCAACAGGCGCGCGCAACAAACAAACACGCTGACGAGAGCGGGGCTTAAAA 4073
QY 22 ValSerAlaAsnThrValThrValAsnGlyArgArgAsnGlnArgArgThrGlyArg 41
DB 4074 CTTCCCGCGTACGTGCGACCGCATP----- 4097
QY 42 GlnValSerProProAspAsnPheThrAlaAlaAlaGlnAspLeuAlaGlnSerLeuAsp 61
DB 4098 ---ACGACCGCGCGAGATGGCCGACCCGAC-----CAC 4130
QY 62 AlaAsnThrValThrPheProAlaAsnIleSerSerMetProGluPheArgAsnTrpAla 81
DB 4131 GCGAACAACCGGGGTCAA-CCGGGCTG-----GACACGCGTGAAGGGGCT 4174
QY 82 LysGlyLysIleAspLeuAsp-----SerAspSerIleGlyTyrPheLysTyr 98
DB 4175 AAGGCAAGCATGATGAGAGTCCCATGCGACCACTCAAGGCGATTCACGACTAT 4234
QY 99 LeuAspProAlaGlyValThrGluSerAlaArgAlaValAlaGlyLysTyrSerLysIlePro 118
DB 4235 CTAGACCCCGGAGCGAGATCAACAGACGACCTGACACAGGG-----AAATTCCTCC 4285
QY 119 AspGlyLeuValLysPheSerValAspAlaGluIleArgGluIleTyrAsnGluGlyCys 138
DB 4286 GACGGCGCATCTCAGTCAATCGATGCGGTCAATTCGAGGACCGTGGCGCGACGATAC 4345
QY 139 ProValValThrAspValSerValProLeuAspGlyArgGlnTyrPheSerIlePhe 158
DB 4346 CCGGAGCTGAATTCCTACAGCGCTACCGCTGATGCGGAGCGGACCTGCTACTAGTATG 4405
QY 159 SerPheProMetPheArgThrAlaTyrValAlaValAlaAsnValGluAsnLysGluMet 178
DB 4406 CATCTCCGTTCTTCAAGCATCCGTTGTTCATCAACACCAACACACAGCAAGAA--- 4462
QY 179 SerLeuAspValValAsn---AspLeu-----IleGluTyrPheAsnAsnLeu 193
DB 4463 ---GTCAAGAGACGACGCGCATCTGATGCTGTCGCGCAACGATGG---AACAAACAG 4516

```


QY	62	AlAsnThrValThrPheProAlaAsnIleSerSerMetProGluPheArgAsnTrpAla	81
Db	2406	GCAGAACCCGGGCTCAA-CCGCGGTGC-----GACAGCGTGAAGGGGCT	23633
QY	82	LysGlyLysIleAspLeuAsp-----SerAspSerIleGlyTrpTrpPheLysTrp	98
Db	2362	AAGCGAACGATATATGAGATGATCCCATGGCAGCAGCATCAAGCATGATTCACGACAT	23033
QY	99	LeuAspProAlaGlyAlaThrGluSerAlaAlaGlyValGlyGluTrpSerLysIlePro	118
Db	2302	CTAACACCCGAGCAGAGATACAGACAGACCTGACAGCGGG-----AAATATCC	22523
QY	119	AspGlyLeuValLysPheSerValaAspAlaGluIleArgGluIleTrpAsnGluLeuLys	138
Db	2251	GACGGCCCGATACCTAGTCATACATCCGGCTCATTTTCGAGGACCGTGCCGCCAGATAC	21922
QY	139	ProValValThrAspValSerValProLeuAspGlyArgGluTrpSerLeuSerIlePhe	158
Db	2191	CCGGAGCATGAATTCATCGACCGCTACCGTGATGGGGGACCTGGCCCTCTACTAGTATG	21322
QY	159	SerPheProMetPheArgThrAlaTrpValAlaValAlaAsnValGluAsnLysGluMet	178
Db	2131	CATCTCCGGTCTTCAGGCATCCGTTGTGTTCATCACCCACACAGCAACAGGAA--	20757
QY	179	SerLeuAspValValAsn--AspLeu-----IleGluTrpLeuAsnLeu	193
Db	2074	---GTCAAGATGACGAACGCCGATCTCGATGGCTTCGACAGAGATTTGG---AACACAGG	2021
QY	194	AlaAspTrpArgTrpValValaAspSerGluGlnTrpIleAsnPheThrAsnAspThrThr	213
Db	2020	ACGGAGCTGGACCCGAAAGCGAGC---TACCCAGTTGGGGCCAAAGTCGGAGAC-----GTG	19707
QY	214	TyrTrpValArgIleArgValLeuArgProThr-----TyrAspValProAspPro	230
Db	1969	TTTTTACATGGTCGTC-----CCGACCGAAGCGCTACGAGACGTACACACCCCG	19222
QY	231	ThrGlu--GlyLeuValArgThrValSerAspTyrArgLeuThrTyrLysAlaIleThr	249
Db	1921	ACTCACTACGGGTGATATCAGGCTTACCTCGAGAGTTACCGTCTGACATCAGCGCGCTCACA	18623
QY	250	CysGluAlaAsnMetProThrLeuValaAspGlnGlyPheTrpIleGlyValGlyTrpAla	269
Db	1861	GCGACTCTCAACGACCCACTCTCTGTAATCAAGGAGTGGCGGTGTCGCGCATGTC--	18053
QY	270	LeuThrProThrSerLeuProGlnTyrAspValSerGluAlaTrpAlaLeuHisThrLeu	289
Db	1804	---CAACCGGCACAAAGAACACAGAAAGGAACCCGCAATGATGTAGCGGGACACCCAA	17488
QY	290	Thr-----PheAlaArgProSerSerAlaAlaIleuAlaPhe	302
Db	1747	ACGGGGCGGAACGTACAGCTCGGCGGGTGTGAGGCGGCATACATGACATTCGACGTGAC--	16913
QY	303	ValTrpAlaGlyLeuProGlnGlyValThrAla--ProAlaGlyTrpProAlaTrpGlu	321
Db	1690	ATCGGGGACACAGGTCAAGTTCGGGGGGCGACAGCATCCCTTACCCAGCGTGTCAATGGGG	16313
QY	322	GlnAlaSerSerGlyLysTrpLeuThrTrpArgHisAsnGlyValThrThrPheProAlaGly	341
Db	1630	CCGATGCCGAGAGTGGGGAGCTGTGTGTCCAGACTGCACATCTGACATTCACAGCTCGGA	15717
QY	342	-----SerValSerTrpValLeuProGlnGlyPheAlaLeuGluAlaArgTrpAspPro	358
Db	1570	AACACAAATCACCAATCAGACACACGCTGCCACACAGGGTGGTG-----	15293
QY	359	AsnAspGlySerTrpThrAspPheAlaSerAlaGly--AspTrpValThrPheArgGln	377
Db	1528	---ACGGGAATGTGGCAATTCACAGCCACAGCGGAGGACACACGCTGAC-----	14813
QY	378	ValAlaValaAspGluValValValThrAsnAsnProAlaGlyGlyGlySerAlaProThr	397
Db	1480	---GTGGAC-----GCCGGAGCG-----	14663

OY	398	PheThrValArgValIProProSerSerSnaLarYrThraSnTrValPheArgASnThLeu	417
Db	1465	--ACTGATTCGGTTCGGAGCGAATTGTGACGCCCTGCAGACTGAATCTGCAGCAATCAA	1409
OY	418	LeuglunrArqProSerSerArqrArgleuGluePromeProProAlaAspPhegly	437
Db	1408	CtCAATCAGAATTCACCACCAAC-----GACATGMAACCCAAATGATCGACGC	1364
OY	438	GlnThrValAlaAsnAsnProLySlleGlueInSerLeuLeuys-GluThrLeuGlyCy	457
Db	1363	-----AACGCCAAGACCATTCAGTTCCAACTACGAAMAAGAGGACCATTAATG	1316
OY	457	sTryLeuValHisSerLyMetArgAsnProValPheGlnLeuthrProAlaSerSepH	477
Db	1315	CCTG-----AGCCGTCAAATCCGTGTTCMAATGACAAT-GGACGCTCTTA	1272
OY	477	eGIValaValserPheAsnAsnProGLYTyrGlnArgthrArgasp----LeuProasPT	496
Db	1271	TGGACCGGTCCAT-----GAACACCCAGGACGAACATGTGTAGATT	1230
OY	496	yrrHrgly-----lleargaspSerPheaspGlnAsnMetSer-----	508
Db	1229	AACCCGGGAATTTGTGTGCACCTCCAAGATATACATGCAGACAACTTCCGATFAGTTGCC	1170
OY	509	-----ThralaValalaHis-PhearGerLeuSerHisSerCysSerIlleValThr	525
Db	1169	gtcGAtgAcCGGTATGTCTACATCAACCGTACCCTACTTGCAAGGTGTTCCGACGCTTCG	1110
OY	526	LysThrTyrgln-GlyTTPgluGlyValThrasnValasnThrPropheGlyGlnPheAl	545
Db	1109	AMGCATACCGCGCGAGGAG-----CCTTGGGGCCCCCTTGC	1071
OY	545	aHislaaglyLeuLeuLysAsnGluGlnLeucCysLeuAlaAspAspleuAlaThrAr	565
Db	1070	TAGTCGCACACCTCCGAAGAGCACGTGGCGCTACAGTGGCTGCACCTTGACGATCT	1011
OY	565	gleuthrGlyValTyrrProAlaThrAspAsnPhelaAlaValaserAlaPheAlaI	585
Db	1010	GCACCATTCGCATCCCGGAACGATACACGATTCGGGCGCCATTCCGGATGGTGGC	951
OY	585	aAsnMetLeuSerSerValleuLysSerGlnLarThrSerSerIlleLysSerValGl	605
Db	950	CNAGACCAATAGCCCAATRCCCTGC-----TATGTGCATCA-----	914
OY	605	yGluThrAlaValGlyAlaAlaGlnSerGlyDeuAlalysLeuProGlyLeuLeuMetSe	625
Db	913	-----GCAGCGGAGTAGCGAATGCGGTGACGAGCTAGCATAGAGACGGACCGAGAG	861
OY	625	rValProGlyLysIleAlaAlaValaValaArgAlaAlaArgAlaArg-----	641
Db	860	TTTACCTTCGAAATTCACCTCGAAGAGCGCGCAACGAAGAGGACGACCTGTGGCGGAAT	801
OY	642	-----ArgAlaAlaArg 645	
Db	800	CGCTCGAGAGGCCCGC 785	
RESULT 15	AE005131	11659 bp DNA linear BCT 12-FEB-2001	
LOCUS	AE005131		
DEFINITION	Halo bacterium sp. NRC-1 section 162 of 170 of the complete genome.		
ACCESSION	AE005131 AE004437		
VERSION	AE005131.1 GI:10581942		
KEYWORDS			
SOURCE			
ORGANISM	Halo bacterium sp. NRC-1. Halobacterium sp. NRC-1 Archaea; Euryarchaeota; Halobacteria; Halobacteriales; Halobacteriaceae; Halobacterium. 1 (bases 1 to 11659) Ng,W.Y., Kennedy,S.P., Mahites,G.G., Bergquist,B., Pan,M., Shukla,H.D., Lasky,S.R., Baliga,N., Thorsson,Y., Shrogon,J., Swartzell,S., Weir,D., Hall,U.J., Dahl,T.A., Welti,R., Goo,Y.A., Lettichsen,B., Keller,K., Cruz,R., Hanson,M.J., Hough,D.W., Maddock,S.D.G., Jablonowski,P.E., Krebs,M.P., Angevine,C.M., Dale,H.,		
REFERENCE			
AUTHORS			

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

Isenbarger, T. A., Peck, R. F., Pohlschrod, M., Spudich, J. L.,
Jung, K.-H., Alam, M., Freitas, T., Hou, S., Daniels, C. J., Dennis, P. P.,
Omer, A. D., Ebhardt, H., Lowe, T. M., Liang, P., Riley, M., Hood, L. and
DasSarma, S.
Genome sequence of *Halobacterium* species NRC-1
Proc. Natl. Acad. Sci. U.S.A. 97 (22), 12176-12181 (2000)
20504483
11016950

2 (bases 1 to 11659)
Ng, M. V., Kennedy, S. P., Mahairas, G. G., Bergquist, B., Pan, M.,
Shukla, H. D., Lasky, S. R., Baliga, N., Thorsson, V., Sorogna, J.,
Swartzell, S., Weir, D., Hall, J., Dahl, T. A., Welti, R., Goo, Y. A.,
Leithauser, B., Keller, K., Cruz, R., Danson, M. J., Hough, D. W.,
Maddocks, D. G., Jablonski, P. E., Krebs, M. P., Angevine, C. M., Dale, H.,
Isenbarger, T. A., Peck, R. F., Pohlschrod, M., Spudich, J. L.,
Jung, K.-H., Alam, M., Freitas, T., Hou, S., Daniels, C. J., Dennis, P. P.,
Omer, A. D., Ebhardt, H., Lowe, T. M., Liang, P., Riley, M., Hood, L. and
DasSarma, S.
Direct Submission
Submitted (14-JUL-2000) Institute for Systems Biology, 4225
Roosevelt Way NE, Seattle, WA 98105, USA

FEATURES
source
location/Qualifiers
1..11659
/organism="Halobacterium sp. NRC-1"
/strain="NRC-1"
/db_xref="taxon:64091"
157..1074
/gene="ygeC"
/note="VNG2553G"
157..1074
/gene="ygeC"
/note="ygeC"
/codon_start=1
/transl_table=11
/product="6-phosphogluconate dehydrogenase"
/protein_id="AAG20606.1"
/db_xref="GI:10581943"
/translation="MELGVGLGRGRIVVRCDCDAGHDVVAFDIDATARDADAGA
TPADTLDAELAEIGKRIWLPAGDPDADELAPLSDVYVVGSGNSHEDST
BRAEHDPAVLDGSGSGPAGAEISLWVGSPKAYVDIYVPDPAVATGPDGHRMG
PAGAGHYKYNHNGEYVNLKOTYGGFELLGGRIDLEAVARTWNNGANVIRSHLE
LCEAFREBSDLGTADVDHAGVGSTGTWVEALEQVEPLVVSQALSERGSRSTGS
DTGFRARLRNLRYGFRHNEVARGDA"
complement(1071..1652)
/gene="VNG2554H"
complement(1071..1652)
/gene="VNG2554H"
/note="VNG2554H"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="Vng2554h"
/protein_id="AAG20607.1"
/db_xref="GI:10581944"
/translation="MDVAVLLFEGFDELDATIGPVEVFTAAELGDDLDVSKRTLVPE
GVTARHGIVAPDVLFGVPDLVVPGGWTDKRPQVWSEVENETLPERLDTLHGGS
ATTASVCTGAMLADAGLLDGRPAATHASADDLAEVDVDRERYDDGTVLTAAGY
TAGIDLALHVERECSAGLAEOVAETLEYASTY"
1746..2324
/gene="VNG2555C"
1746..2324
/gene="VNG2555C"
/note="VNG2555C"
/note="conserved hypothetical protein"
/codon_start=1
/transl_table=11
/product="Vng2555c"
/protein_id="AAG20608.1"
/db_xref="GI:10581945"
/translation="MWTTLGVTLGLGVTLFVVLHLAKGSSKPPAEDIADQVLEHRAS
TVPEPTPEPVPNRNIGGGAGAVAGAESELEDGEEDDGFDDAIADDDVEYEIE
YKGETIYVANSNLSILDAGVDEGMDLPYACQGGQCVSCSGRVADGEDAAEYIHNHN
ESTGDAIEEGYMLTCTAHPASSTIETDERP"
2413..2485
/gene="trn44"
/gene="trn44"

gene
CDS

tRNA
gene
CDS

2413..2485
/gene="trn44"
/product="tRNA-Arg"
complement(2545..2955)
/gene="VNG2556H"
complement(2545..2955)
/gene="VNG2556H"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="Vng2556h"
/protein_id="AAG20609.1"
/db_xref="GI:10581946"
/translation="MTVVDRCGLLIAGVGLNLAWMTLHSPITATAADSGGLGVL
GIMALVIGPMLGAIPTNHWADHCAGVCPILATGVLTCACLMNGITAPPSLSLTVY
EAMRFVLVYVVGAECCITRGHVNDSNRSSQ"
complement(3099..4019)
/gene="fepC"
/note="VNG2558G"
complement(3099..4019)
/gene="fepC"
/note="fepC"
/codon_start=1
/transl_table=11
/product="ferric enterobactin transport protein"
/protein_id="AAG20610.1"
/db_xref="GI:10581947"
/translation="MTROPADATGTDPDGLSGTDLVYRPGDDPYIDGESYTP
GDTVALIGPNSGSKSTLNLANKNSPADGTLLDGHITDELKSKELRAGLSLGSQ
NVTPGDISVRELVERGRYPHGFDFSDADARAAYTALRAVGDVHLGSRVSLSG
OKOLVIMAMALADPTDYLDELPTTFLDPHQLEVMYDIQALRDVHTVVVYVLDIN
OARVADVIALKDGTTIARCGPADVYTKETLRVFDIEATVSTTSHQMVPTEALH
DEDGDHGHARGSRPTRETAADGTPPAE"
complement(4023..5102)
/gene="yfmD2"
/note="VNG2560G"
complement(4023..5102)
/gene="yfmD2"
/note="yfmD2"
/codon_start=1
/transl_table=11
/product="ferrichrome ABC transporter permease protein"
/protein_id="AAG20611.1"
/db_xref="GI:10581948"
/translation="MTANTDASATRRATTAAGSLVVIATLILVSGAYPEPIADA
MHTVDEPDIWSSAGVWRATLLGOLPELSPROYVNNIRLRLNCTINGANLPIADA
AIPQIVTRNEASPIYILIGVNSGAGMYLLITLTSGLLPPMLAAGAYAFILIVA
IAMKNGTNPVRLVLAGVYVSTVFGSVORALFFIDNLGVNSAQTWLSGLLGTDMAQ
VRIALPTVLSLALAFVLTNLDVLLGEEFAESLGMPEVKVRFVAGVAILSTAAAI
AVAGLIGFVGLIVPHMVNLTIGNNTRDLAGCVALIGPALLVGAADVARGARLADPVOLPY
GITGLVGSPYFLXMRNTENLGD1"
complement(5109..6263)
/gene="VNG2562H"
complement(5109..6263)
/gene="VNG2562H"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="Vng2562h"
/protein_id="AAG20612.1"
/db_xref="GI:10581949"
/translation="MARRROIAGASLIASLAGCTSSQSPPTDGSSTTGTAGSDSY
TVVEPNPPTFESVPEYTAIIPVMMDIGALGIIQPSATSLGRAPLKFYDALPDVA
FHKAEIROLASDSESGYDKENFYAADDDVHLIDPRMLKPYANWTDDDDIIAIESAGP
MGSSTIRRASSAAGDGPFDYLYEAFEKRAELFORARFEMQSKKALMNTISGLRP
ESDRPTVAALIRGVRPDTGPRFAPVLPDDQNTLITRLGKQDDFAFAGVDPGPIGTGE
LLVDVDPYIGAVGGLTSQTHAEFVDVVEPEANPNQQLTAAVNGMLVPTGGQFMGP
IVDLSTETALAKWYYPDAFGSPGSPPEIAPSERLFDORVADAINGVGL"
complement(6353..7216)
/gene="VNG2563H"
complement(6353..7216)
/gene="VNG2563H"
/note="hypothetical protein"

gene
CDS

OY	368	rfaGlyaApThrVal---	ThrPheArgGlnValAlaValaAspGluValValaValaThrAsp	387
Db	1482	GTGGGGCCACCGTGGAGCCCGTGGCGCGGCTC-		1515
OY	387	nAsnProAlaGlyGlySerAlaProThrPheThrValArgValProPheSerAsnAl		407
Db	1516	-ACGCTTGCTGGGGACAGAGCGTGGGCATCGACACACTGAGTGGCGGCGCCACTCTCCG		1574
OY	407	aTyThrAsnThValAlaPheArgAsnThrLeuLeu-		421
Db	1575	G-----GCCGTGGGGAACACTCTGTACGGGGCGATGGCTCCAGTTCTGTGAA		1622
OY	421	gProSerSerArgArg-LeuGlnLeuProMetLeuProProAlaAspPheGlyGlnThrVala		441
Db	1623	CCCTTGAAACAGGAGACGTGCAACGTTCATACCCTGCTGCAC-		1666
OY	441	laAsnAsnProLysIleGluGlnSerLeuLeuLysGluThrLeuGlyCysTyrrLeuValH		461
Db	1667	--CCACACCCCTCA-		1678
OY	461	lAserLysMetArgAsnProValAlaPheGlnLeuThrProAlaSerSerPheGlyAlaValS		481
Db	1678			1678
OY	481	ePheAsnAsnProGlyTyrrGluArgThrArgAspLeuProAspTyrrThnGlyLeArga		501
Db	1679	----AATTGGCCCGGA-		1691
OY	501	sPSePheAspGlnAsnMetSerThAlaValAlaHAsnPheArgSerLeuSerHIsSerC		521
Db	1692	AAACGTGGTGGCGCATTCACAAATACGTGTAACACTGGCGTTCACATACCGCTACAGCATGT		1751
OY	521	ySerLLeuValThrLysThrTyrrGlnGlyTrrpGluGlyValThrAsnValAsnThrPro		541
Db	1752	ACCACACTGGGTCTCACGCTCGCGCTGGG----		1804
OY	541	heGlyInPheAlaHAsIaGlyLeuLeu-Lys-----AsnGluGlnIleLeuCysLeu		558
Db	1805	-----GCATCTCGCGAAGGGCTATCCACACCCCGCGGAGCAATC-----		1847
OY	559	AlaAspAspLeuAlaThrArgLeuThrGlyValTyrrProAlaThrAsp-AsnPhe-----		576
Db	1848	GCACAGAGAGCTCTCTGCACACCCGTGCTCCACAGTGCCTGGAGACGAGCACTTCCCTGAGCCG		1907
OY	577	-----AlaAlaAlaValSerAlaPheAlaAlaAsnMe		587
Db	1908	TACAAACCGCTCATCGGGGCGGGCGGGGCGGGCGGTCTGCTGGCGGTGCAGACTCG		1967
OY	587	tleuSerSerValLeuLysSerGlnAlaThrSerSerIleIleLysSerValGlyGluTh		607
Db	1968	GAACTCTGAAGACGGCGAGGAGGAGACGACGGCTTGATCTCGGAGCGATCGCCGACGAC		2027
OY	607	rAlaValGlyAlaAlaGlnSerClyLeuAlaLysLeuProGlyLeuLeuMetSerValPr		627
Db	2028	GACGTCTGACTACTAGCAGATCGAGTACGTCAAG-----		2061
OY	627	oGlyLysIleAlaAlaArgValAlaArgAlaArgAlaArgAlaAlaAlaArgAlaAs		647
Db	2062	-----AAGCGAGACCATCATGAGTGTCCGAGCAACGAGAGCATC		2099
OY	647	n 647		
Db	2100	C 2100		